

UNITED STATES DISTRICT COURT  
EASTERN DISTRICT OF WISCONSIN

---

BRENDA KOEHLER, KELLY PARKER,  
LAYLA BOLTEN, and GREGORY HANDLOSER,

Plaintiffs,

Case No. 13-cv-885-pp

v.

INFOSYS TECHNOLOGIES LIMITED INC.,  
and INFOSYS PUBLIC SERVICES INC.,

Defendants.

---

**ORDER GRANTING DEFENDANT’S MOTION TO EXCLUDE EXPERT  
OPINIONS OF DAVID NEUMARK (DKT. NO. 97), DENYING WITHOUT  
PREJUDICE PLAINTIFFS’ MOTION FOR PARTIAL SUMMARY JUDGMENT  
(DKT. NO. 86), DENYING WITHOUT PREJUDICE PLAINTIFFS’ MOTION TO  
CERTIFY CLASS (DKT. NO. 88), DENYING AS MOOT JOINT MOTION FOR  
HEARING (DKT. NO. 202) AND SETTING STATUS CONFERENCE**

---

The defendants—Infosys Technologies Limited, Inc. and its wholly-owned subsidiary Infosys Public Services, Inc. (collectively Infosys)—are an international IT company headquartered in India. The plaintiffs are Caucasian American nationals who claim they either were not hired, were not promoted, or were fired based on the defendants’ “systematic pattern and practice of discriminating against non-South Asian employees and in many instances replacing them with South Asian employees.” Dkt. No. 19 at ¶¶6, 96. The plaintiffs moved for class certification, and the parties filed cross-motions for summary judgment. After the plaintiffs filed their motion for partial summary judgment and their class certification motion, the defendant moved to exclude

the opinions of the plaintiff's expert witness, Dr. David Neumark, who analyzed the defendant's demographic data relative to the plaintiffs' claims that Indians and South Asians were overrepresented in the defendant's employee population and were favored over non-Indians and non-South Asians in terms of hiring, promotion and termination. Dkt. No. 97.

Because Dr. Neumark was not qualified to perform analyses to determine employees' national origins, and because he used unreliable methodologies to do so, the court will grant the motion to exclude his opinion.

## **I. Facts and Procedural History**

### **A. The Plaintiffs' Allegations**

In its order denying the defendants' motion to dismiss the second amended complaint, the court described the plaintiffs' allegations:

The four plaintiffs allege that they are Caucasian individuals of American national origin, against whom [Infosys] made adverse hiring or employment decisions on the basis of their race and national original. They seek to represent a class of similarly situated individuals. The complaint alleges that [the two entities that make up Infosys] are corporations organized and headquartered in India. The corporations have many offices located in the United States that are comprised predominately—or in some cases, entirely—of employees of the South Asian race and of Indian, Bangladeshi, and Nepalese national original. (Dkt. No. 19, ¶22.) The complaint contains allegations describing the defendants' discriminatory treatment of each of the plaintiffs, and describing the defendants' alleged employment practices that cause a disparate impact against Caucasians.

\* \* \* \* \*

In addition to the plaintiffs' individual experiences, the complaint contains allegations regarding the defendants' purported intent to discriminate on the basis of race. The complaint alleges that two [Infosys] executives, who are South Asian, explicitly encouraged recruiters to focus their efforts on recruiting Indian

candidates and dismissed complaints that highly qualified American candidates were being rejected in favor of Indian candidates. Id. ¶¶28-30. The complaint further alleges that an [Infosys] hiring manager stated: “There does exist an element of discrimination. We are advised to hire Indians because they will work off the clock without murmur and they can always be transferred across the nation without hesitation unlike [a] local workforce.” Id., ¶3.

The plaintiffs allege that the defendants achieve their discriminatory objectives, at least in part, by their practice of setting annual “visa quotas” to support the growth of their United States offices, hiring South Asian workers in sufficient numbers to meet those quotas, and securing visas for foreign workers to enter and work in the U.S. Id., ¶¶31-35. They allege that [Infosys] sets annual growth targets for its U.S. offices, then budgets for the number (and expense) of additional foreign-worker visas [Infosys] will need to secure in order to bring foreign workers into the U.S. to meet its targets. Id., ¶33. The complaint alleges that this practice results in annual visa quotas of South Asian workers to be hired and “employed within the U.S., irrespective of the fact that qualified workers exist in the U.S. that Infosys could use to support its U.S. business.” Id.

Dkt. No. 31 at 2-3, 4-5.

The court<sup>1</sup> concluded that the second amended complaint stated “causes of action for disparate treatment and disparate impact on the basis of race and national origin in violation of Title VII, and for disparate treatment on the basis of race in violation of §1981.” Id. at 17.

#### B. Discovery

Discovery disputes began shortly after the scheduling conference and the court’s issuance of the scheduling order. Three months after issuing the

---

<sup>1</sup> The case originally was assigned to Judge Charles N. Clevert, Jr. It was reassigned to this district judge in late December 2014, after the undersigned was appointed to the district court. By that time, the parties had briefed the motion to dismiss the second amended complaint; the undersigned ruled on that motion.

scheduling order, the court held a status conference at which the plaintiffs argued that the defendants were not willing to turn over information regarding employees working in the United States on visas, or information about the defendants' visa policies. Dkt. No. 38 at 1. The defense responded that they had data from two databases containing hundreds of thousands of records and containing information about employees hired in India and employees hired in the U.S. Id. The plaintiff asserted that the defendants were providing information on "base hires," and was defining that term in a way that excluded the information the plaintiffs needed. Id. The defense responded that it could not figure out what information the plaintiffs needed. Id. at 2. Eventually the court ordered the defendants to turn over information they indicated they had compiled and asked the plaintiffs to advise the court in writing of the amount of time they'd need to review the information. Id. Less than a month later, the plaintiffs requested a discovery hearing (in a fifty-six-page filing that included six exhibits). Dkt. No. 43. The parties briefed this request, and the court held another hearing. Dkt. No. 47. At the hearing, the plaintiffs alleged that the defendants were trying to control what they turned over and in what form, while the defense agreed that the data it had provided needed interpretation and suggested that perhaps bringing in a magistrate judge to assist in the exchange of data might help. Id. at 1-2. The court agreed to refer the discovery issues to a magistrate judge for mediation. Id. at 2.

The parties worked with Magistrate Judge Nancy Joseph for about four months—unsuccessfully; Judge Joseph reported that mediation had not

“resulted in resolution or settlement of all the discovery issues in this case.” Dkt. No. 55. This court held another status conference. Dkt. No. 57. The plaintiffs’ counsel listed four categories of information the plaintiffs still sought; defense counsel indicated that the parties might be able to work through those issues if they had more time to talk. Id. at 2. The court scheduled another status conference. Id. at 3. It also referred the case to Magistrate Judge David Jones, hoping that he could assist the parties in working through the discovery issues.

Judge Jones worked with the parties regularly—often weekly—for over two years, stopping only when he left the court.

C. The Defendants’ Motion to Exclude Expert Opinions

The plaintiffs filed a motion for partial summary judgment. Dkt. Nos. 77, 86. The defendants filed their own motion for summary judgment. Dkt. No. 78. The following day, the plaintiffs filed a motion under Federal Rule of Civil Procedure 23 for certification of three classes. Dkt. Nos. 81, 88.<sup>2</sup> The motion defined the classes as follows:

A. Hiring Class: All individuals who are not of South Asian race or Indian national origin who sought a position with Infosys in the United States and were not hired from August 1, 2009 through the date of class certification.

---

<sup>2</sup> The plaintiffs filed the motion for partial summary judgment and attachments, dkt. no. 77, as well as the motion for class certification and attachments, dkt. no. 81, as “restricted” from public view, but also filed objections to the defendants’ designation of many of the attachments as confidential, dkt. no. 76. Three weeks later, the plaintiffs notified the court that the defendants and a third party had agreed to withdraw certain confidentiality designations, and that the plaintiffs would refile the motion for partial summary judgment and the motion for class certification. Dkt. No. 85.

B. Promotion Class: All individuals who are not of South Asian race or Indian national origin who were employed by Infosys in the United States between August 1, 2009 and the date of class certification for a period of at least 18 months and were not promoted.

C. Termination Class: All individuals who are not of South Asian race or Indian national origin who were employed by Infosys in the United States between August 1, 2009 and the date of class certification and were terminated.

Dkt. No. 88 at 1.

The plaintiffs attached as an exhibit to both their motion for partial summary judgment and the motion for class certification a September 2016 expert report from Dr. David Neumark. Dkt. Nos. 77-2 (attachment to motion for partial summary judgment); 88-2 (attachment to motion for class certification). Neumark, a professor of economics at the University of California-Irvine, described himself as “a labor economist who ha[d] done extensive research on labor market discrimination, including methods for measuring and testing for discrimination that have been adopted by many other researchers.” Id. at 4, ¶1. Neumark stated that the plaintiffs had hired him “as a statistical expert to evaluate claims of discrimination at Infosys Technologies Limited, Inc. . . . with respect to its hiring, promotions, and terminations in the United States.” Id. at 4, ¶3. Specifically, Neumark said that the plaintiffs had asked him “to evaluate whether the data are consistent with discrimination against applicants and employees at Infosys who were non-South Asian or non-Indian (and, correspondingly, who were white, black, Hispanic, or other categories not in the South Asian or Indian groups).” Id. at 5, ¶3.

The defendants simultaneously filed a motion to exclude Neumark's opinion, dkt. no. 97, their brief in opposition to the plaintiffs' motion for partial summary judgment, dkt. no. 99, and their brief in opposition to the plaintiffs' motion to certify classes, dkt. no. 103. The defendants' brief in support of the motion to exclude Neumark's opinion asserted that the plaintiffs' motion for partial summary judgment and their motion for conditional class certification "depend entirely on the opinions of their statistics expert, Dr. David Neumark." Dkt. No. 98 at 8. The defendants argued that the court should strike Neumark's opinions and exclude them from the case because his opinions were based on "pseudoscience." Id. at 9.

The plaintiffs filed a brief in opposition to that motion. Dkt. No. 123. They argued that Neumark is a "highly-qualified professor who performed straightforward and reliable statistical analyses" and that his statistical findings "are relevant to proving Plaintiffs' *prima facie* case under the [*International Brotherhood of*] *Teamsters* [*v. United States*, 431 U.S. 324 (1977)] framework." Id. at 5. The plaintiffs explained that after seeing the defendants' "attacks" on Neumark's methodologies, Neumark "produced a short, supplemental report, addressing, among other things," what the plaintiffs characterized as "ancillary issues" raised by the defendants. Id. at 9. The plaintiffs attached to this motion a February 2017 supplemental expert report from Neumark. Dkt. No. 123-7.

The defendants replied that faced with their critique of Neumark's methodologies, the plaintiffs had shifted from asserting that Neumark's

statistics proved discrimination to arguing that Neumark had offered no opinion on causation and had opined only that the disparities he observed could not have occurred by chance. Dkt. No. 128 at 3.

On June 11, 2020, the court heard argument on the motion.<sup>3</sup> Dkt. Nos. 195 (audio recording of hearing), 196 (minutes), 204 (transcript). After advising the parties that it did not agree with either of their interpretations of Adams v. Ameritech Services, Inc., 231 F.3d 414 (7th Cir. 2000), dkt. no. 204 at 15, line 17 through 19, line 20, the court heard argument regarding Neumark's methodologies and analyses, dkt. no. 204 at 20, line 2 through 64, line 12. The court took the motion under advisement.

#### D. The Other Pending Motions

As indicated, the defendants filed a motion for summary judgment. Dkt. No. 78. The parties have fully briefed that motion—the plaintiffs have filed their opposition materials, dkt. no. 93, and the defendants have filed a reply in support, dkt. no. 108.

A little over three weeks after the defendants filed their motion for summary judgment, the plaintiffs filed their motion for partial summary judgment. Dkt. No. 86. The parties have fully briefed that motion—the defendants have filed their opposition materials, dkt. no. 99, and the plaintiffs have filed a reply in support, dkt. no. 117.

---

<sup>3</sup> The court ruled on other motions at that hearing—a motion to join a plaintiff (Dkt. No. 157) and a motion to strike (Dkt. No. 177). Dkt. No. 204 at 7, lines 20-23. The court denied the motion to join a plaintiff. Id. 9, lines 19-20. It granted the motion to strike and asked the defendant to resubmit the documents that were the target of that motion. Id. at 12, lines 20-25.



The parties have more than fully briefed the plaintiffs' motion for Rule 23 class certification. Dkt. No. 88. The defendants filed opposition materials, dkt. no. 103, and the plaintiffs filed a reply, dkt. no. 116. The defendants then filed a sur-reply,<sup>4</sup> dkt. no. 125, and the plaintiffs filed what the court supposes is a sur-sur-reply,<sup>5</sup> dkt. no. 132.

Because of the court's extreme delay in addressing the motion to exclude Neumark's opinions, the parties filed a motion asking the court to schedule a telephonic status conference to discuss whether there was anything they could do to "facilitate the Court's consideration of the pending matters." Dkt. No. 202.

## **II. The Legal Posture of the Claims**

Because the plaintiffs filed their motion for partial summary judgment and their class certification motion at the same time, and because the defendants based their motion to exclude Neumark's opinions on the arguments in the plaintiffs' motions, the parties' arguments on the motion to exclude Neumark's opinions often are intertwined with the substance of the plaintiffs' claims and the defendants' arguments in opposition to those claims. The parties reference their pleadings on the summary judgment and class certification motions when arguing the motion to exclude, and the plaintiffs

---

<sup>4</sup> Neither the federal rules nor this court's local rules allow sur-replies, but it appears that Magistrate Judge Jones allowed the defendants to do so. See Dkt. No. 120 at 2.

<sup>5</sup> The plaintiffs sought leave to file this document, dkt. no. 129, and Judge Jones granted it, dkt. no. 131 at 1-2.

sometimes conflate legal standards for class certification and for their burden of proving their *prima facie* case of discrimination with the law governing the admissibility of expert witness testimony. It is helpful (for the court, at any rate) to understand the plaintiffs' claims and the law governing them before turning to the admissibility of Neumark's opinions.

There are different types of employment discrimination. "Disparate treatment . . . is the most easily understood type of discrimination." Teamsters, 431 U.S. at 335 n.15. Disparate treatment occurs when "[t]he employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin." Id. When a plaintiff alleges disparate treatment, "[p]roof of discriminatory motive is critical, although it can in some situations be inferred from the mere fact of differences in treatment." Id. Another type of employment discrimination is "disparate impact"— "employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity." Id. "Proof of discriminatory motive . . . is not required under a disparate-impact theory." Id. A third type of employment discrimination claim is a "pattern or practice" claim. "Pattern-or-practice claims, like [disparate treatment] claims, represent a theory of intentional discrimination." Puffer v. Allstate Ins. Co., 675 F.3d 709, 716 (7th Cir. 2012) (citing Council 31, Am. Fed'n of State, Cnty. & Mun. Emps, AFL-CIO v. Ward, 978 F.2d 373, 378 (7th Cir. 1992)). A pattern or practice claim "require[s] a 'showing that an employer regularly and purposefully

discriminates against a protected group.” Id. A plaintiff bringing a pattern-or-practice claim is required to prove “that discrimination ‘was the company’s standard operating procedure—the regular rather than the unusual practice.’” Id. (quoting Teamsters, 431 U.S. at 336).

The plaintiffs have alleged both disparate treatment, dkt. no. 19-2 at ¶¶130-139, and disparate impact, id. ¶¶140-142, and in their disparate treatment claims, they have alleged a pattern or practice of discrimination, id. at ¶¶130-139. To prevail on those claims at trial, the plaintiffs must prove by a preponderance of the evidence that the defendants engaged in intentional, purposeful discrimination and that that discrimination was their standard operating procedure.

“[T]he general principle [is] that any Title VII plaintiff must carry the initial burden of offering evidence adequate to create an inference that an employment decision was based on a discriminatory criterion illegal under” Title VII. Teamsters, 431 U.S. at 358. A plaintiff may do this in a few ways. An individual plaintiff seeking to prove intentional employment discrimination may use the burden-shifting framework articulated in McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973), a “helpful way to evaluate evidence of discriminatory intent in employment discrimination claims.” Brooks v. Avancez, 39 F.4th 424, 433-34 (7th Cir. 2022). For a plaintiff in a pattern-or-practice case, the

initial burden is to demonstrate that unlawful discrimination has been a regular procedure or policy followed by an employer or group of employers . . . . At the initial, “liability” stage of a pattern-or-practice suit the [plaintiff] is not required to offer evidence that each

person for whom [the plaintiff] will ultimately seek relief was a victim of the employer's discriminatory policy. [The plaintiff's] burden is to establish a *prima facie* case that such a policy existed. The burden then shifts to the employer to defeat the *prima facie* showing of a pattern or practice by demonstrating that the [plaintiff's] proof is either inaccurate or insignificant.

Teamsters, 431 U.S. at 360.

So—having alleged pattern-or-practice claims, the plaintiffs first must prove a *prima facie* case, which requires them to prove by a preponderance of the evidence that the defendants had a system-wide policy or practice of discrimination. The plaintiffs' motion for partial summary judgment asserts that they have proven that *prima facie* case, based on the statistical disparities revealed in Neumark's work. Dkt. No. 86. They ask the court to find, as a matter of law, that they have proven the *prima facie* case, such that the burden must shift to the defendants to rebut the inference of discrimination at trial. Id. at 6. In their motion for class certification, the plaintiffs assert that they have satisfied all the elements of Fed. R. Civ. P. 23 for class certification, including showing that they have presented common evidence of discrimination in hiring, promotion and termination; some of the evidence they cite is Neumark's work. Dkt. No. 88.

In all the pending motions—the plaintiffs' motion for partial summary judgment, the defendants' motion for summary judgment, the plaintiffs' motion for class certification and the defendants' motion to exclude Neumark's opinions—the parties discuss the role of statistics in proving a *prima facie* case of discrimination in a pattern-or-practice case. They often talk past each other. The defendants assert, for example, that Neumark's opinions are not relevant

because they will not assist the trier of fact in deciding whether the defendants engaged in a pattern or practice of discrimination. The plaintiffs respond that Neumark's opinions "are relevant in proving Plaintiffs' *prima facie* case under the *Teamsters* framework." Dkt. No. 123 at 5. The defendants' argument, while couched in terms of relevance, really challenges the reliability of Neumark's assumptions and the methodologies he applies to test them. The plaintiffs' argument states the obvious.

The Seventh Circuit repeatedly has held that statistics play a significant role in proving a *prima facie* case of discrimination in a pattern-or-practice case. See, e.g., Adams, 231 F.3d at 424 ("There is no presumption that statistical evidence has no useful role to play in disparate treatment employment discrimination cases—indeed, we are hard pressed to see how anyone could take such a position consistently with the Supreme Court's guidance on the matter and this court has not done so."); E.E.O.C. v. O&G Spring & Wire Forms Specialty Co., 38 F.3d 872, 876 (7th Cir. 1994) ("Appropriate statistical evidence can . . . be sufficient to establish a pattern and practice of discrimination . . . ."); E.E.O.C. v. Chi. Miniature Lamp Works, 947 F.2d 292, 297 (7th Cir. 1991) (stating that a *prima facie* case for a pattern or practice of disparate treatment can be established by statistical evidence showing "substantial disparities," shored up by evidence of general policies or specific incidents); Shidaker v. Tisch, 833 F.2d 627, 630 (7th Cir. 1986) (disparate impact case, but noting that "in order to make out a *prima facie* case of discrimination under a disparate treatment 'pattern and practice' theory, a

plaintiff must compare the percentage of minorities or women in the employer's workforce" with the people in the labor force who possess the relevant qualifications); Coates v. Johnson & Johnson, 756 F.2d 524, 532 (7th Cir. 1985) (in a pattern-or-practice case, "[t]he plaintiffs' prima facie case will thus usually consist of statistical evidence demonstrating substantial disparities in the application of employment actions as to minorities and the unprotected group, buttressed by evidence of general policies or specific instances of discrimination").

In their motion for partial summary judgment, the plaintiffs go further, explaining that the statistical analyses (Neumark's analyses) show such a gross disparity between the demographics of the defendants' hires, promotions and terminations and those in the relevant labor market that, standing alone, the statistics prove the plaintiffs' *prima facie* case. In support of this claim, the plaintiffs correctly state that the Supreme Court and the Seventh Circuit have stated that in the appropriate case, statistics alone can prove a *prima facie* case of discrimination in a pattern-or-practice case. The Supreme Court has held that "[w]here gross statistical disparities can be shown, they alone may in a proper case constitute prima facie proof of a pattern or practice of discrimination." Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 307-08 (1977) (citing Teamsters, 431 U.S. at 339). Likewise, the Seventh Circuit has explained that "[r]eliance on statistical evidence by no means diminishes the plaintiff's obligation to prove discriminatory intent—but in some cases, statistical disparities alone may prove intent." O&G Spring, 38 F.3d at 876.

But the issue the defendants have raised in the motion that is the subject of this order is not whether statistics play an important role in pattern-and-practice cases, or whether it is possible for statistics alone to suffice to prove a plaintiff's *prima facie* case in a pattern-and-practice case. The issue is whether *Dr. Neumark's* statistics are the kind of "appropriate statistical evidence" that suffice to prove—if not standing alone, then buttressed by other evidence—the plaintiffs' *prima facie* case that the defendants engaged in a pattern or practice of discrimination sufficient to allow the court to grant them partial summary judgment, or that the plaintiffs have met the Rule 23 standard for proceeding as a class action. "[S]tatistics are not irrefutable; they come in infinite variety and, like any other kind of evidence, they may be rebutted. In short, their usefulness depends on all of the surrounding facts and circumstances." Teamsters, 431 U.S. at 340. "[S]trong statistics may prove a case on their own, while shaky statistics may be insufficient unless accompanied by additional evidence." O&G Spring, 38 F.3d at 876 (citing Teamsters, 431 U.S. at 340; Chi. Miniature, 947 F.2d at 300-01). Statistics have limited value; they "can only show a relationship between an employer's decisions and the affected employees' traits; they do not show causation." Radue v. Kimberly-Clark Corp., 219 F.3d 612, 616 (7th Cir. 2000). Given the role statistics play in pattern-or-practice cases, the soundness of the statistical analyses matters. "The pattern-or-practice case starts with a stronger showing than the individual disparate treatment case; the 'prima facie case' under *McDonnell Douglas Corp. . . .*, supports only a weak inference of discrimination,

while the statistical showing in a pattern-or-practice case leaves a smaller possibility of race-neutral conduct.” Mister v. Ill. Cent. Gulf. R.R. Co., 832 F.2d 1427, 1434 (7th Cir. 1987). And when a plaintiff relies on “statistically significant” disparities revealed by its statistical evidence, one must keep in mind that “the finer points of significance-testing are pertinent only if the analyst has formulated the hypothesis correctly and decided what pattern (if established) will confirm or refute that hypothesis.” Id. at 1431.

The parties each have rooted through the Seventh Circuit’s discrimination cases to find fact patterns that approximate the facts in this case. Many of those cases are not helpful because the lower courts addressed the statistical evidence in the context of summary judgment or at trial. The parties sifted through decisions from the Seventh Circuit and other courts to find sentences that support their respective positions; given the many cases on the subject, each was able to find sound bites that, taken in isolation, would mandate a ruling in their favor. Both parties succumbed to the temptation to try their cases; the plaintiffs repeatedly emphasized the dramatic disparities Neumark’s analyses showed, while the defendants previewed how they might decimate his conclusions on cross-examination.

In most respects, these arguments were beside the point. In considering a challenge to the admissibility of an expert witness’ opinions or testimony, the court’s role is to ensure that the expert provides “[p]ertinent evidence based on scientifically valid principles.” Daubert v. Merrell Dow Pharm., Inc., 509 U.S.



579, 597 (1993). The court concludes that Dr. Neumark’s opinions and analyses do not fit this bill.

### **III. Dr. Neumark’s Opinions**

#### **A. The September 2016 Report (Dkt. No. 88-2)**

##### **1. *Qualifications***

In the September 2016 report the plaintiffs attached to their motion for class certification, Dr. Neumark—a professor of economics at UC-Irvine and a labor economist who has “done extensive research on labor market discrimination, including methods for measuring and testing for discrimination” that he indicates “have been adopted by many other researchers”—indicated that he had published approximately twenty-five peer-reviewed journal publications “on discrimination based on race, ethnicity, sex, or age.” Dkt. No. 88-2 at 4. He indicated that he had published studies in edited books and published his own book “on sex discrimination and sex differences in labor markets (based on [his own] papers).” *Id.* He explained that the goal of his research was “to better understand the role of discrimination versus other explanations of differences in labor market outcomes by race, ethnicity, sex, or age.” *Id.*

Neumark recounted that he had held positions “at the Federal Reserve Board, the University of Pennsylvania, Michigan State University, and the Public Policy Institute of California.” *Id.* He was, at the time of the report, a research associate with the National Bureau of Economic Research and a research fellow at the Institute for Study of Labor in Germany. *Id.* He indicated

that at UC-Irvine, he directed “the Economic Self-Sufficiency Policy Research Institute.” Id.

Neumark explained that the plaintiffs were paying him \$450 per hour for his work. Id. at 5. At the end of his introduction, he provided the following qualifier:

It is important to note that this analysis is based on my best current understanding of the data we have on Infosys applicants and employees, and the other information with which I have been provided. It is possible that I will learn more about the Infosys data, company procedures, and other matters in the course of this case, which could lead to changes in the analysis and findings.

Id.

## 2. *Summary of Findings*

Neumark’s September 2016 report is thirty-three pages long with forty-six pages of appendices. At the outset, he provided a summary of his findings:

6. First, analysis of Infosys and external benchmark data are strongly consistent with Infosys discriminating in favor of South Asians (and Indians) in hiring employees placed in United States-based positions. For example, from 2009 through 2015, 89.39% of Infosys’ United States workforce was South Asian while only 11.45% of the United States’ Computer Systems Design and Related Services industry was South Asian. Thus, the share of South Asian workers in Infosys’ United States-based workforce, when compared to the relevant labor market, is 301.17 standard deviations higher, and the statistical likelihood that this disparity is due to chance—as opposed to a systematic difference in hiring favoring one group over the other—is less than 0.0000001%, or less than 1 in 1 billion (Table 1). When limiting the analysis or the relevant labor market to only those positions within the United States’ Computer Systems Design and Related Services industry that are most common at Infosys, the share of South Asian workers is 20.28%, compared to 89.39% at Infosys. Based on this comparison, the share of South Asian workers in Infosys’ United States-based workforce, when compared to the relevant labor market, is 201.88 standard deviations higher, and the statistical likelihood that this disparity is due to chance is less than 0.0000001%, or less than 1 in 1 billion. When doing the same

analysis for Indians and non-Indians in the Computer Systems Design and Related Services industry, and the positions most common at Infosys, the numbers are similarly stark. Respectively the standard deviations are 302.16 and 202.46 (Table 3).

7. Further, the data are strongly consistent with Infosys discriminating in shaping its own applicant pool in the United States to favor South Asians and Indians. And from this applicant pool, South Asians and Indians receive job offers at considerably higher rates than non-South Asians and non-Indians.

8. Second, analysis of Infosys' promotions data provides evidence that is strongly consistent with Infosys discriminating in favor of South Asians and Indians in promotions. While 57.74% of Infosys' Indian employees in the United States had job roles that ended in promotions to a higher job level, only 20.85% of its non-Indian employees were promoted to a higher job level—a difference of 53.45 standard deviations (Table 12). The statistical likelihood of obtain[ing] either of these results by chance is less than 0.0000001%, or less than 1 in 1 billion.

9. Third, analysis of Infosys' termination data provides evidence strongly consistent with Infosys discriminating against non-South Asians and non-Indians in termination decisions. Among all Indians whose employment with Infosys ended, 11.4% were asked by Infosys to leave, while among all non-Indians whose Infosys employment ended, 22.8% were asked to leave—a standard deviation difference of 11.57. The statistical likelihood that this result was obtained by chance is 0.0000001%, or less than 1 in 1 billion (Table 15).

10. To summarize, in the period covered by the complaint, Infosys employed 46,979 workers in 134,113 roles in the United States. Compared to the relevant labor market, the Infosys workforce was composed of a remarkably disproportionate share of South Asians, and similarly a remarkably disproportionate share of Indians (a slightly narrower group than South Asians). Across the different analyses this report covers, these percentages were approximately 89%, versus about 11%-20% in the relevant labor market. Infosys hires were also disproportionately South Asian or Indian to a remarkable degree—approximately 75%, versus approximately 11%-20% of employees in the relevant labor market. The applicants to Infosys were also highly disproportionately South Asian or Indian. Despite an employment share of about 11%-20% South Asian or Indian in the relevant labor market, Infosys' applicants were around 45% South Asian or Indian. And even when it received applicants who were not South Asian or Indian, Infosys was more likely to offer

jobs to applicants who were South Asian or Indian. The apparent favorable treatment of South Asians or Indians continued once employment commenced. Promotion rates for South Asian or Indian employees were about 58%, compared to around 21% for non-South Asian or non-Indian employees. Separation rates of non-South Asian or non-Indian employees were much higher—approximately 51% versus 8%—and in particular the rate of company-initiated separations was double for non-South Asians or non-Indians—about 23% versus 11% for South Asians or Indians. Along all of these dimensions I have examined, the statistical evidence is strongly consistent with Infosys discriminating against non-South Asians and non-Indians in hiring, recruitment, promotions, and terminations.

Id. at 5-7.

### 3. *Methodology*

Neumark explained that he used a three-step analysis. Id. at 7. First, he indicated, it was “necessary to define the relevant labor market.” Id. Second, it was necessary “to determine what Infosys’ workforce would look like absent discrimination in favor of South Asians (and Indians);” he called this the “null hypothesis.” Id. Finally, he stated that it was necessary to “compare the null hypothesis with the actual composition of Infosys’ United States-based workforce, and determine the statistical strength of the magnitude of disparity between the two.” Id. at 7-8. He measured this disparity in “standard deviations”—“the ratio of the estimate of the difference between the observed representation of South Asians (or Indians) and the representation under the null hypothesis (computed as the difference between what is observed at Infosys, and what is observed in the ACS<sup>6</sup> data), divided by the standard

---

<sup>6</sup> Later in the report, Neumark explained that “ACS” is the acronym for the American Community Survey. Dkt. No. 88-2 at 9.

deviation of that estimated difference.” Id. at 8 n.2. He explained that “the more standard deviations from the null hypothesis the representation of South Asians (or Indians) at Infosys is, the less likely it is that a given result was due to chance, as opposed to a systematic difference in hiring,” and that “[a] disparity of two standard deviations is generally sufficient to show that a result is extremely unlikely (less than a 5% probability) to be caused by chance.” Id. at 8.

At step one, Neumark defined the “relevant labor market” as “the market for United States’ computing workforce.” Id. at 9. He said he defined that market broadly—“simply the industry—specifically, the Computer Systems Design and Related Services Industry in which Infosys operates”—and more narrowly—“the occupations within this industry that are the three most common occupations at Infosys: Computer and Information Systems Managers . . . , Computer Systems Analysts . . . , and Software Developers, applications and systems software . . . .” Id. He explained that using these three occupations to define the market was “appropriate” because they made up 98% of Infosys’ U.S.-based workforce, thus providing “external benchmarks” against which to compare Infosys’s own workforce; he also asserted that it was appropriate to compare Infosys’ hires to its own applicant pool, and that these occupations provided an “internal benchmark” by which to do so. Id.

Neumark obtained his data on the U.S. computer workforce from the American Community Survey for the years 2009-2014. Id. He explained that the ACS provided “a large and detailed sample of the United States population,”

surveying 1% of the U.S. population annually, and he asserted that the sample sizes “are large enough to provide accurate estimates of the composition of the workforce by industry and occupation, for large industry and occupation categories.” Id. He asserted that the ACS contained “detailed information on industry and occupation of employment, and on race, ethnicity, and birthplace.” Id. He explained, however, that while both the ACS data and the Infosys data used the racial classification “Asian,” neither used the classification “South Asian.” Id. at 9 n.4. Neumark considered the ACS data only for adults aged 18 to 65. Id. at 9. Neumark also identified the Infosys-provided data he’d used for the analysis, explaining that a “new hire” was anyone with a date of hire after January 1, 2009. Id. at 10. He obtained data on *applicants*, as opposed to hires, from the two software systems Infosys uses to manage applications—one called “SAP” and one called “Kenexa.” Id. The SAP data covered May 2010 through August 2013; the Kenexa data covered April 2013 through August 2015. Id.

Neumark then blended steps two and three of the analysis—comparing the “null hypothesis,” or the percentages reflected in the relevant labor market under his broad and narrow definitions, with the percentages at Infosys. He started with the workforce and new hires. Using the broad definition of the labor market—the market for the United States’ computing workforce—Neumark stated that the ACS data showed that the U.S. computer workforce consisted of 11.45% South Asians and 88.55% non-South Asians. Id. He explained that under the null hypothesis “of no discrimination based on

ethnicity,” he would expect the Infosys United States-based workforce “to consist of approximately the same percentages of South Asian and non-South Asian workers as in the relevant labor market.” Id. He said he’d also expect Infosys to hire for positions in the U.S. at about the same percentages of South Asians and non-South Asians as the relevant labor market. Id. at 10-11. Neumark reports, however, that his analysis showed that Infosys’ U.S.-based workforce consisted of 89.39% South Asians and only 10.61% non-South Asians; he stated that the difference between the percentages at Infosys and the percentages in the broader U.S. market was 301.17 standard deviations. Id. at 11. Using the narrower definition of the relevant market—the three occupations Neumark had identified—Neumark stated that the U.S. workforce consisted of 20.28% South Asian individuals and 79.72% non-South Asians, 201.88 standard deviations from the Infosys percentages. Id. He stated that between January 1, 2009 and the end of 2015, 75.18% of Infosys’ U.S. hires were South Asian, while the U.S. computer workforce consisted of 11.45% South Asians, a difference of 194.81 standard deviations. Id.

In a footnote, Neumark stated the following:

South Asian ethnicity was determined using birthplace from [a particular data set provided by Infosys]. South Asians include the following ethnicities: Indian, Sri Lankan, Bangladeshi, Pakistani, Burmese, Nepalese, and Afghan. If an Infosys employee was born outside of a South Asian country, then the list of South-Asian last names at Infosys was used to determine ethnicity, with potentially non-South-Asian names recoded as non-South-Asian.

Id. at 13, n.8. Similar footnotes appear throughout the report for different parts of the analysis.

Neumark conducted the same analyses regarding national origin, comparing the national origins (particularly, Indian) of Infosys' workforce with the broad and narrow definitions of the relevant labor market and finding standard deviations in the hundreds. Id. at 14-16.

Neumark also discussed "the external benchmarks and hiring of only U.S. citizens," stating, "[t]hat is, deputees are excluded." Id. at 18. In stating that the "share of Indians among Infosys hires of U.S. citizens [was] higher than in the external benchmarks from the ACS," Neumark dropped a footnote that said, "There are only 5 non-Indian South Asian Americans [sic] hires in the Infosys data. Due to the small sample size of applicants who are South Asian but not Indian, the results are virtually identical to the analysis using Indian applicants." Id. at 18 n.12. This information appears in other footnotes in other sections of the report. Neumark concluded that for the broader U.S. market, "the share Indian in the relevant labor market is 11.25%, and the share non-Indian is 88.75%," while "among U.S. citizens hired at Infosys, 26.98% are Indian, and 73.02% non-Indian," 28.69 standard deviations. Id. at 18. For the narrower market, "the percent Indian in the relevant labor market is 20.02% Indian and 79.98% non-Indian, compared to 26.98% Indian and 73.02% non-Indian among Infosys U.S. citizen hires," 9.78 standard deviations. Id. at 18-19.

Neumark then included a section titled "Shaping the Applicant Pool and Hiring from the Applicant Pool." Id. at 21. He explained that Infosys kept data on "applications by United States-based applicants for United States-based



positions.” Id. He said that under the null hypothesis, he would expect Infosys’ applicant pool in the United States to closely mirror the relevant labor market. Id. He concluded that the ACS data showed that 11.53% “of the United States’ Computer Systems Design and Related Services industry (all positions) is Indian, while the share of Infosys’ applicants that are Indian is 45.09%,” 183.94 standard deviations. Id. For the narrower definition of the relevant market, 20.45% “[were] Indian—again, compared to 45.09% of Infosys applicants.” Id. In this part of the analysis, Neumark said he had reviewed deposition transcripts and declarations, which led him to understand “that the disparity in Infosys’ applicant pool is apparently due at least in part to deliberate efforts to include South Asians and Indians and to exclude non-South Asians and non-Indians.” Id. at 22 n.15. He cited to deposition testimony and declarations from Samuel Marrero, a former recruiter with Infosys who worked in its Plano, Texas office starting in October 2011, see dkt. no. 94-44, and the declaration of Davina Linguist, an Infosys recruiter who alleged that she had not received a raise or promotion when others of South Asian race or Indian national origin had, see dkt. no. 94-5.

Neumark stated that “[e]ven accepting the flawed applicant pool,” he would expect under the null hypothesis that South Asian and Indian applicants would “receive offers of employment at approximately equal rates as non-South Asian and non-Indian applicants.” Id. at 23. He indicated that the Kenexa data (covering applicants between April 2013 and August 2015) showed that Infosys made written offers to 4.14% of Indian applicants versus 3.58% of

non-Indian applicants, 5.67 standard deviations higher. Id. He stated that he reviewed the SAP data—covering applicants from May 2010 through August 2013—and found that Infosys made offers to 7.72% of Indian applicants versus 6.12% of non-Indian applicants, 8.63 standard deviations higher. Id. at 23-24. He stated that when he combined the Kenexa and SAP data, Infosys made offers to 5.28% of Indian applicants versus 4.44% of non-Indian applicants, 9.25 standard deviations higher. Id. at 24.

Having analyzed Infosys' hires, Neumark applied his three-step analysis to Infosys promotions. He explained that Infosys recorded employees' job history; each job (or role) that an employee had was "recorded with their job title, a beginning date, and an end date." Id. at 26. The data also included the job level, the "work stream in which it was located," and the employee's termination date. Id. Neumark explained how many employees and how many job roles were included in this data, noting that on average, an employee held 2.9 roles. Id. He defined "promotions" as "a vertical move upward in job levels." Id. at 27. He explained how he'd coded for the job levels, based on the two ways Infosys had identified them during the relevant period. Id. at 27-28. He stated that "each change in role" was considered a "unique observation," and stated that he'd calculated that there was a "55.40% promotion rate when a role is changed." Id. at 28.

Turning to the definition of the relevant labor market, Neumark defined that market as "Infosys' workforce in the United States from August 1, 2009 to 2015," and asserted that it was an appropriate market "because promotions

occur only for those already employed at Infosys.” Id. at 29. He then turned to the comparison of the null hypothesis with the data, and concluded that “[o]f all Indian employees who were employed at some point between 2009 and 2005, 57.74% were promoted when switching roles,” compared to 20.85% “[f]or non-Indian employees,” 53.45 standard deviations. Id.

Finally, Neumark applied his three-step analysis to employee terminations. He stated that the Infosys data codes indicated that 16.44% of employment separations were due to “Company Initiative,” which he interpreted to mean involuntary termination or discharge. Id. at 30. Neumark opined that it was “likely . . . that this number understates the true number of involuntary terminations, as employees may be pressured to resign or may be given the option of resigning instead of being fired, and not coded as ‘Company Initiative.’” Id. at 30-31.

He again defined the relevant labor market as Infosys’ workforce in the U.S. between 2009 and 2015, “because separations of employment can only occur among Infosys employees.” Id. Comparing the relevant workforce with the null hypothesis, Neumark concluded that “[o]f all Indian employees who were employed at some point between August 1, 2009 and 2015, 7.55% had separations of employment (3,169 employees),” while for non-Indian employees, “the separation rate was 50.64% (2,532 employees),” 88.21 standard deviations. Id. at 31. Looking specifically at separations coded as “Company Initiative,” Neumark stated that 22.79% of non-Indians separated at “Company

Initiative,” compared to 11.36% of Indians, 11.57 standard deviations. Id. at 32.

Neumark attached a nine-page appendix explaining how he matched Infosys and ACS data on “occupational structure.” Id. at 34-42. He also provided a four-page appendix titled “Data on Ethnicity/National Origin.” Id. at 43-46. This appendix explained that while ACS race data was coded into “broad categories of Asian, white, black, Hispanic, and other,” it also provided “classifications of national origin, which allow[ed] [him] to identify Indians and South Asians.” Id. at 43. Infosys data, however, coded employees as “white, Asian, Hispanic/Latino, black/African-American, American Indian, Native Hawaiian/Pacific Islander, or two or more races.” Id. The Infosys data coded race for 86% of employees; “the remainder do not report in which case they are excluded from the analysis by race/ethnicity.” Id. Neumark explained:

3. The race, ethnicity classifications in the Infosys data do not differentiate between someone of Indian or Southeast Asian descent and someone of, for example, Chinese descent; both groups would be coded as Asian. There is information on which employees were born in India or other countries from [records provided by Infosys]. For these records, of course, it is straightforward to identify South Asians or, more narrowly, Indians. The challenge is identifying South Asians or Indians among applicants, and among employees not born in South Asian countries or in India.

4. To attempt to identify employees and applicants in these groups, I use data on Infosys employees born in South Asian countries or India (separately). In particular, I identify all surnames from the employee file . . . that are associated with an employee who was born in a South Asian country or in India, and use these names to identify South Asians or Indians among records on employees not born in those countries, or among records on applicants. This implies that I match 100% of South Asian or Indian employees born in a South Asian country or in India. However, it is possible to falsely identify South Asians or Indians among Infosys employees not born

in India, and among applicants. This is not likely to happen if a name appears in the Infosys data for an employee born in a South Asian country or in India, but it is a more Western name.

5. To reduce the changes of a false identification, I attempted to identify non-South Asian or non-Indian names that appear in the list of surnames associated with employees born in South Asian or in India. There are 21,093 unique surnames among Infosys employees. To look for misleading names, I restrict my search to surnames that are associated with two or more employees, and for which there is the potential for misclassification because there is at least one employee born in a South Asian country or in India, and at least one employee not born in a South Asian country or in India with that name. This led to a list of 1,273 surnames. Reading through this list, I identified names that do not appear to be Indian—specifically, identifying 63 surnames that potentially identify large numbers of non-South Asians or non-Indians and could lead to false identifications in the data. These surnames, along with their frequency at Infosys and the share of employees with that surname who were born in a South Asian country or in India, appear in Appendix Table 2.1. Note that the percentages are very high. But that is because, as documented below, South Asians and Indians are so strongly over-represented at Infosys.

Id. at 43-44. In reference to the sixty-three surnames that, in Neumark’s view, could lead to false identification, Neumark stated in a footnote, “These are names that clearly appear to be non-South Asian/non-Indian. To be sure, this list could be refined further, but the effects are likely to be extremely negligible.” Id. at 44 n.40.

Appendix Table 2.1 is titled “Non-Indian Surnames Held by Indian Employees at Infosys.” Id. at 45. The table is divided into three columns, titled “Surname,” “Share Indian” and “Frequency.” Id. The highest “share Indian” percentage is 95%—the names “Mathew” and “Sebastian.” Id. The highest “frequency” name is “Thomas.” Id. Several names appear with 50% “Share Indian” and 2 “Frequency”—Andrew, Benjamin, Carney, Clement, Dawson,

Gabriel, Grace, Newton, Norris, Raymond, Richard, Rubin, Souza, Stanley and Stone. Id. at 45-46.

In sum, Neumark considered an Infosys *employee* to be Indian or South Asian if (a) the employee was born in India or South Asia, or (b) the employee's last name matched the last name of another Infosys employee born in India or South Asia and was not one of the sixty-three "Western" names removed from the list. He considered an *applicant* to be Indian or South Asian if the applicant's last name matched the last name of an Infosys *employee* born in India or South Asia and was not one of the sixty-three "Western-sounding" names removed from the list. If an employee or applicant did not meet one of the preceding criteria, that employee or applicant was considered non-Indian or non-South Asian.

B. The February 2017 Report (Dkt. No. 123-7)

In the February 2017 report Neumark prepared after seeing the defendants' critiques of his methodology, Neumark first said he'd "done the external benchmarks analysis state by state, based on the location at which Infosys employees spent the most time in the United States." Dkt. No. 123-7 at 3. He stated that he'd rank-ordered the observations by the number of Infosys employees, "so it is easy to look at the evidence for the states where Infosys has the largest presence." Id. He provided a chart that listed the state, the "Share of South Asian ACS respondents within Computer Systems Design and Related Service industry," the "Total ACS respondents within Computer Systems Design and Related Services industry," the "Share of South Asian Infosys

Employees,” the “Total Infosys Employees” and the “Standard deviations.” Id. at 4. (The chart does not include all fifty states; Neumark explained that he showed states only if “the number of observations multiplied by the shares South Asian and non-South Asian in both samples both exceed 5, so that the large-sample statistic applie[d].” Id.)

Next, as to promotions, Neumark said he’d modified his analysis “to exclude promotions that occurred outside the United States,” and that he had included only individuals “employed at Infosys for a minimum of 18 months, consistent with the class definition.” Id. at 5. He also “corrected for a coding issue that had resulted in a number of incorrectly classified promotions as a result of Infosys’ September 2009 transition from letter job levels to number job levels.” Id. And he provided a table “that looks only at promotions after September 2009 to avoid this transition issue altogether.” Id.

For terminations, Neumark compared the rate of “Company Initiative” involuntary terminations “for base employees, relative to all base employees, for terminations occurring in the United States.” Id. at 8. In a footnote, he indicates that when he prepared the initial report, he “did not know how to categorize employees as base or deputies,” and that he was “unaware of data that distinguished between terminations occurring in the United States versus those that occurred elsewhere;” he indicated that he had corrected for these issues. Id. at 8 n.5.

He explained why he’d considered only written job offers and not verbal ones, in response to a criticism by the defendants’ expert. Id. at 10.

Finally, he wrote:

10. [The defendant's expert] claims that my analysis of job offers reaches the wrong conclusion because my method of classifying applicants as non-South Asian based on names is flawed. In particular, he notes, correctly, that I do not require a last name in the applicant file to match the name of an employee born outside of the South Asian countries to classify them as non-South Asian. But [the defendants' expert] admits there is no perfect way to identify race using names. . . . However, there is a logic to the method I used. In particular, because Infosys employs relatively few non-South Asian employees, the list of non-South Asian names on which to match is quite restricted. It is also important to note that [the defendants' expert] significantly overstates the severity of the classification problem. In his Appendix B, he tabulates names of people who self-identify as Asian, and rank-orders the names in terms of their frequency for those who self-identify as Asian. Note that what he has done here, it appears, is two things: First, [he] also does matching to non-South Asian names to classify non-South Asians the same way I classified South Asians. He then lists the names of those not matched who self-identify as Asian. Then, in the third column, he lists the number with these names who do *not* self-identify as Asian. Unsurprisingly, since he has rank-ordered and selected these based on self-identification as Asian, the latter numbers are small.

11. In fact, it appears that my classification of applicants as South Asian or non-South Asian appears to work rather well. In Table 4.1, I show the rank-orderings by name frequency of applicants classified as non-South Asian, as in my report. This table suggests that the most common names of those classified as non-South Asian are indeed non-South Asian. Although "Manthri" does show up as the most common name (for which everyone self-identifies as Asian), after this, most of the names do not appear to be South Asian, even though many self-identify as Asian.

Id. at 10-11.

There follows a table (Table 4.1) titled "Unmatched or Matched Names in Applicant File, Classified as Non-South Asian, Top 50, Rank-Ordered by Frequency." Id. at 12. It lists columns title Last name, Total applicants, "Self-identify as Asian" and "Unmatched to employees born outside SA." Id. It shows



that there were 339 total applicants with the last name “Manthri,” all of whom self-identified as “Asian;” the “Unmatched to employees born outside SA” column indicates “Unmatched.” Id. A number of other names were unmatched despite having numeric significance in the “Total applicants” column and having most or all of the applicants with that name self-identifying as “Asian”—“bharadia,” “araveeti,” “m,” “v,” “venkatramanan,” periyasamy,” “sinharoy” and “maitreyan.” Id.

#### **IV. The Parties’ Arguments**

##### **A. The Defendants’ Opening Brief (Dkt. No. 98)**

The defendants first challenge Neumark’s assertion that he compared the racial composition of Infosys’ workforce to that of the relevant labor market and his conclusion that the disparities he observed between racial groups could not have happened by chance. Dkt. No. 98 at 10. The defendants assert that Neumark aggregated any employee who worked in the United States for any time between August 1, 2009 and the end of 2015—both in and outside of India. Id. They maintain that Neumark ignored that “81% of the Infosys workforce population is composed of ‘Deputies’—employees hired *in India*<sup>7</sup> and transferred to work in the United States on specific jobs for temporary periods of time.”<sup>8</sup> Id. The defendants report that at his deposition, Neumark could not

---

<sup>7</sup> In every instance where quoted language from the defendants’ brief is italicized, the emphasis is in the original. The defendants make generous use of italics.

<sup>8</sup> The defendants frequently reference and rely on their own expert, Bernard Siskin, Ph.D., director of specialty consulting firm BLDS, LLC and a former

“distinguish Deputies from ‘Base Employees,’ *i.e.*, employees hired in the United States to perform work in the U.S.” Id. at 11. They maintain that “[t]his labor market error alone explains the disparity observed by Dr. Neumark.” Id. They further argue that none of the studies Neumark conducted of the defendants’ data “examine or control for a single explanatory variable that could account for the racial disparities that” he reported. Id.

The defendants then turn to Neumark’s analysis of applicants, asserting that he “systematically mis-classifie[d] the race of job applicants.” Id. They explain that “because Infosys does not determine whether job applicants are South Asian (or Indian)<sup>9</sup>, Dr. Neumark invents a ‘name matching’ method, notwithstanding his complete lack of any expertise in such methodologies.” Id. The defendants point out that Neumark “classifies as South Asian only those applicants whose surname matches that of a South Asian person *in Infosys’s employee database* (and only if the applicant’s surname does not sound non-South Asian to Dr. Neumark).” Id. at 11-12. They argue that applicants without matching surnames—“regardless of their self-identified race or South Asian names”—are not classified as South Asian, even though “*a majority* (54%) of the 245,972 total applications have a surname without any match.” Id. at 12. The

---

tenured faculty member and chair of the Department of Statistics at Temple University in Philadelphia. See Dkt. No. 97-3.

<sup>9</sup> The defendants assert that this is “not surprising since Infosys’ self-reported race categories follow those required by the OFCCP. ‘South Asian’ (or Indian) is not a recognized category.” Id. at 11 n.4. OFCCP is the U.S. Department of Labor’s Office of Federal Contract Compliance Programs. <https://www.dol.gov/agencies/ofccp>.

defendants maintain that rather than disregarding applicants with non-matching names, or finding some way to classify them, Neumark “mistakenly assumes that all un-identified applicants are non-South Asian,” and they argue that this error renders any analysis that relies on his name-matching methodology unreliable. Id.

The defendants also assert that Neumark ignored the racial self-identification that Infosys collects for some applicants, arguing that he could have used it to test the reliability of his “novel methodology.” Id. They explain that “self-reported race is generally available” in the Kenexa software system. Id. at 12 n.6. They say that 14% of the applicants “who self-reported a race *other than Asian* are classified by Dr. Neumark as Indian,” and that “43% of employees who report their race and who Dr. Neumark classifies as non-South Asian self-identify as Asian.” Id.

The defendants argue that Neumark failed to account for the “obvious effect of applicant preferences, including the ‘greater preference by South Asians [relative to non-South Asians] in the U.S. labor market for an Indian employer.’” Id. at 12. They contend that Neumark failed “to account for minimum job qualifications, applicant job history, applicant preferences, or even *who actually applied for any particular job*, meaning his disparities do not reflect different treatment of *similarly situated* people.” Id. at 12-13.

As for promotions, the defendants argue that Neumark did not compare “eligible, similarly situated employees who did and did not receive promotions.” Id. at 13. They assert that Neumark instead relied “on a single data file . . . that

contains the ‘role’ history . . . for every Infosys employee who worked in the United States between August 1, 2009, and the end of 2015, *including roles those employees held overseas—i.e., roles and promotions with no conceivable relevance to this U.S. litigation.*” Id. The defendants characterize Neumark as having “gauge[d] the relative size of two groups of South Asians—employees who changed roles *with* promotions and employees who changed roles *without* promotions—and then compare[d] *that* facially meaningless ratio to the proportion of *non*-South Asians who changed role with and without promotions.” Id. They assert that Neumark analyzed the promotions of South Asian employees “*in India (i.e., promotions of Deputies before they transferred to the U.S.) to conclude that Infosys discriminated when promoting South Asians in the United States,*” maintaining that “*over half of the promotions [Neumark] observes in his promotions study occurred in India.*” Id. at 13-14. They also note that the plaintiffs defined the promotions class as “non-South Asian employees not promoted *after 18 months of working in the United States,*” but that Neumark did not account for this time limitation. Id. at 14.

The defendants allege that in analyzing terminations, one of Neumark’s analyses included voluntary terminations and that he classified as not terminated “the thousands of Deputies whose employment ended after the termination of their United States work and return to India (reflecting Infosys’ standard practice of transferring Deputies to India before formally terminating them).” Id. As to Neumark’s analysis of company-initiated terminations, the defendants assert that Neumark “compares the relative size of two groups of

South Asians—employees who leave Infosys voluntarily and involuntarily—and measures that arbitrary ratio against another one: the proportion of *non*-South Asians who left voluntarily versus involuntarily.” Id. The defendants argue that this approach is illogical:

Assume, for example, that a company employs 10,000 Indian and 10,000 non-Indian employees, and terminates (at the company’s initiative) 100 Indian and 100 non-Indian employees. At the same time, 900 Indian employees decide to leave voluntarily, while only 400 non-Indian employees do so. Though Indian and non-Indian employees are terminated at precisely the same rate, Dr. Neumark would report that Infosys discharged only 10% of the departing Indian employees whereas it discharged 20% of the departing non-Indians—an observation that has nothing to do with discrimination in terminations.

Id. at 14-15.

The defendants assert that Neumark did not consider “any of the obvious explanations (such as performance, tenure, business unit, etc.) for the disparities he observes, despite Dr. Neumark’s access to the necessary information.” Id. at 15.

After discussing these purported flaws, the defendants turn to their legal argument. They begin by asserting that the court should not wait until after it decides the class certification motion to decide whether to exclude Neumark’s opinions. Id. at 15. Citing Am. Honda Motor Co. v. Allen, 600 F.3d 813, 815-16 (7th Cir. 2010), the defendants argue that if an expert’s report is critical to class certification, the court “must” rule on the expert’s qualifications or submissions before ruling on the class certification motion. Id. They maintain that the plaintiffs’ reliance on Neumark’s opinion is critical to class certification and indicate that the plaintiffs have acknowledged as much. Id. at 16.

The defendants then turn to the factors courts must use in evaluating expert testimony—the factors described in, Daubert, 509 U.S. at 589-90, and Federal Rule of Evidence 702. Id. at 16-17. They argue that Neumark’s opinion is not admissible because it is not relevant, linking this to the requirement that an expert’s opinion must assist the trier of fact. Id. at 18. The defendants explain that in a Title VII employment discrimination case, this means “that the evidence must suggest that the protected class was subject to adverse action for discriminatory reasons.” Id. (citing Mozee v. Am. Comm. Marine Serv. Co., 940 F.2d 1036, 1045 (7th Cir. 1991); Radue, 219 F.3d at 616-17). They indicate that in another case, Neumark himself explained that the way an expert typically makes such a showing is “through a regression analysis, in which potential causal variables other than discrimination are studied.” Id. They quote Neumark as saying that the “large body of research on discrimination in labor markets” proceeds by “estimating the effects of membership in a protected group, controlling for other factors that might affect the outcome in question.” Id. at 19.<sup>10</sup> The defendants assert that while, given this, one would have expected Neumark to control for explanatory non-discriminatory factors in this case, he testified at his deposition that there

---

<sup>10</sup> The defendants cite “Neumark Dep. Ex. 13, at 17-18 (citing Rodgers, William M. III, ed., 2006, HANDBOOK ON THE ECONOMICS OF DISCRIMINATION, Cheltenham, UK: Edward Elgar)” as the source of this quote. Dkt. No. 98 at 19. They do not provide a record cite for Neumark’s deposition; it is at Dkt. No. 97-1. Exhibit 13 to that deposition is an expert witness report by Neumark dated September 25, 2008 in Svyerson, et al. v. IBM; the quoted language appears at Dkt. No. 97-1 at 144-45.

could be non-discriminatory explanations that he had not explored. Id. at 19 (citing, *e.g.*, Dkt. No. 97-1 at 63).

The defendants assert that Neumark's admissions that there were non-discriminatory explanations he didn't consider and that he did not try to understand the factors that might underlie promotion and termination decisions are "fatal to any claim of relevance." Id. at 20. They argue that Neumark did not consider any variables other than race—that he did not eliminate any alternative explanations for the disparities he observed and did not control for factors as obvious as, among others, work history or job performance when analyzing termination data. Id. at 20-21. The defendants contend that this flaw renders Neumark's opinions incapable of assisting the trier of fact in determining whether Infosys discriminates in favor of South Asians and therefore renders them inadmissible. Id. at 22.

The defendants also argue that even if they are "incrementally" relevant, Neumark's opinions are unreliable. Id. They reiterate that his opinions are scientifically unreliable because they did not consider any variables other than race or national origin. Id. at 23. They then assert that his opinions are unreliable because Neumark did not base them on "the proper labor market." Id. at 25. The defendants explain:

Dr. Neumark says that the relevant labor market is "the market for United States' computing workforce. Neumark Rpt. 6. But while Dr. Neumark purported to limit his analysis to employees and employment actions in the United States, that is not what he actually did. Rather, Dr. Neumark includes Deputies—employees from India who are temporarily transferred to assignments in the United States—in his workforce snapshot, hiring, promotion, and termination analyses. Admittedly unaware of the difference between

Deputies and Base Employees (Neumark Dep. 57:17-20), Neumark counts a Deputy hired in India as hired domestically. Siskin Rpt. 22 (nothing less than half the “hires” Dr. Neumark includes in his study occurred in the United States). He treats promotions that occurred in India before and after Deputies work in the United States as promotions occurring in the United States for purposes of this case. *Id.* at 51-52 (explaining less than half the promotions Neumark includes in his study occurred in the United States). Yet he treats Deputies who were transferred back to India (and many of whom were terminated there) as employees in the U.S. workforce whose United States-based employment was never terminated. *Id.* at 58. In other words, Dr. Neumark selectively includes Deputies in his studies—and employment actions that occur with respect to them—without any rational basis for doing so or not doing so, and only when the inclusion of this information results in greater disparities in favor of South Asians.

Id. at 26.

The defendants reiterate that 81% of Infosys’ employees working in the U.S. are deputies hired in India and assert that Neumark did not consider the Indian labor market “that is qualified to (and actually does) comprise Infosys’ workforce.” Id. at 26-27. They say that because 81% of Infosys’ workforce “(at any point in time) is from India, even if all of Infosys’ domestic hires during the timeframe Dr. Neumark studies had been non-South Asian, Dr. Neumark would still have observed a statistically significant disparity between the Infosys workforce and the United States labor market.” Id. at 27. Similarly, they argue that Neumark includes positive employment actions (hires and promotions) that occur with respect to deputies while they are working overseas, but ignores adverse employment actions (terminations) that occur overseas. Id. They assert that “in no event can Dr. Neumark rationally maintain that the relevant labor market is limited to just the United States while simultaneously *including* employment decisions that occurred in India to the



extent it helps Plaintiffs' position." Id. at 27-28. And they argue that Neumark provides no justification for limiting the labor market to the United States. Id. at 28.

The defendants say that Neumark's name recognition methodology is flawed and unreliable. Id. at 29. They describe his methodology as "novel," noting that it identifies every individual in Infosys' employee database who says that he or she was born in India or South Asia, assumes that each of them are of South Asian race, develops a list of unique surnames from that group, then compares that list with a list of surnames of all applicants "and assumes all applicants with surnames that match a name on the list as Indian or South Asian—*unless* the applicant [h]as a surname he subjectively identifies to sound 'Western.'" Id. And, they say, Neumark treats "*anyone* with a non-matching name as non-South Asian and non-Indian." Id. The defendants argue that the result is that Neumark "materially inflates the number of unsuccessful applicants who are deemed to be non-South Asian, and, consequently supposedly harmed by not being hired by Infosys." Id.

The defendants use the surname "Manthri" as an example. Id. They explain that the applicant data contained almost 340 applicants with that surname, and that the surname did not match any surname on Neumark's list, so Neumark treated every one of them as non-South Asian and non-Indian, without identifying any basis for doing so. Id. And they argue that he disregarded the self-identification data that Infosys did collect. Id. at 30. They report that in his deposition, Neumark admitted that he did not have any

expertise in name recognition. Id. They assert that Neumark’s approach “ignores the obvious fact that a person’s surname is not necessarily indicative of his national origin,” citing a Fifth Circuit case criticizing the use of “Spanish-surname registration” as problematic. Id. at 30 n.17 (citing Rodriguez v. Bexar Cnty., Tex., 385 F.3d 853, 866 n.18 (5th Cir. 2004)).

The defendants argue that the court cannot determine the error rate for Neumark’s name-matching methodology because it does not have one—the methodology was created for this case. Id. at 30-31. They contend that even without a known error rate, the method is unreliable because Neumark based his list on the names that show up in Infosys’ employee data, not on a census of Indian or South Asian names. Id. at 31. They state that “Dr. Neumark’s apparent conclusion that a list of surnames based on 46,979 employees with 21,093 unique surnames at a single company (many of whom are not Indian or South Asian) reasonably reflects the universe of surnames originating from a country of 1.2 billion people is facially absurd.” Id. They argue that Neumark’s method leads to “irrational results”—Indians who were not hired but did not have a surname that matched the list would be treated as non-Indians who were not hired, employees who self-reported their race as non-Asian were treated as Indian. Id. And the defendants characterize Neumark’s exclusion of people whom he subjectively determined to have Western-sounding surnames as arbitrary. Id. at 32.

Finally, the defendants asserted that Neumark’s individual analyses were based on flawed assumptions. Id. They say his applicant analysis treats all

applicants as if they applied for the same job, rather than factoring in the fact that applicants applied for a variety of jobs in different locations and with different qualification requirements. Id. Accordingly, they argue that he did not compare similarly situated persons. Id. at 33. He did not consider whether the applicants were qualified for the jobs they sought. Id.

The defendants criticize his promotion analysis because Neumark assumed that the only time an employee could be promoted is when that employee's role changed, without any evidence that that was the case. Id. at 34. (They compare an employee whose role draws to its natural conclusion with one who is promoted out of an active role and whose position is backfilled. Id.) They assert that Neumark did not consider the race of the employees who actually were under consideration for promotions, instead comparing the race of employees who entered new job roles but were not competing for promotions with those who entered new job roles because of promotions. Id. at 35.

The defendants criticize Neumark's termination analysis because he disregarded termination decisions that occurred in India. Id. They assert that in looking at "Company Initiative" terminations and assuming them to be involuntary, Neumark did not look at whether Indians were involuntarily terminated at higher rates than non-Indians, "which is the statistic that actually matters." Id. at 36. They point to Neumark's speculation that a downward trend in an employee's ratings could be the result of Infosys manipulating the ratings to justify the termination, asserting that there is no evidence to support this speculation. Id.

B. The Plaintiffs' Opposition Brief (Dkt. No. 123)

The plaintiffs explain that when they asked Infosys to produce “documents related to ‘the demographics or statistics of Infosys’ United States work force,” a discovery dispute ensued. Dkt. No. 123 at 5. They recount that at one hearing, defense counsel told the court that Infosys had a “barn yard” of data to produce. Id. And, the plaintiffs say, Infosys *did* produce a “vast quantity of data” to the plaintiffs. Id. at 6. The plaintiffs explain that the way the data was organized made it difficult to analyze—it was hard to define promotions, difficult to track which promotions happened in the U.S. versus in India, hard to distinguish visa holders from non-visa holders, etc. Id. The plaintiffs recount that the parties had engaged in several meet-and-confers, during which the defendants told the plaintiffs that they did not have any additional documents or data about the demographics or statistics of their U.S. workforce, other than what they’d already produced. Id. The plaintiffs maintain that this representation was false, because as a contractor for the federal government, the defendants are required to “conduct affirmative action analyses.” Id. at 6-7. They argue that to do this, Infosys supplies data to “PeopleFluent,”<sup>11</sup> which analyzes the data and annually provides Infosys with “statistical reports concerning Infosys’ United States workforce and its hiring, promotion, and

---

<sup>11</sup> The plaintiff’s opposition brief does not explain what “PeopleFluent” is. The company’s website says that “[a]s a market leader in integrated talent management and learning solutions, PeopleFluent helps companies hire, develop, and advance a skilled and motivated workforce.” It says that PeopleFluent provides “best-of-breed talent management software and learning solutions that help you realize the full value of your workforce.” <https://www.peoplefluent.com/why-peoplefluent/about-peoplefluent/>.

termination practices—the very issues that are the subject of Dr. Neumark’s analysis.” Id. at 7.

The plaintiffs reiterate Neumark’s qualifications and assert that the defendants haven’t challenged them. Id. They indicate that they retained Neumark “to evaluate the data produced by Infosys in discovery to determine whether the data was consistent with discrimination by Infosys in hiring, promotions, and terminations,” and they recount his conclusions. Id. at 7-8. The plaintiffs assert that many of the defendants’ criticisms of Neumark’s opinions “center on ancillary issues that relate to a subset of Dr. Neumark’s analyses and that stem from the complexity of the data Infosys produced in discovery.” Id. at 8. They note that Neumark produced a “short, supplemental report addressing, among other things, these ancillary issues,” and they argue that Neumark’s supplemental analyses are “consistent with his initial analyses, and continue to demonstrate strong statistical disparities in the hiring, promotion, and termination rates for South Asians versus non-South Asians (as well as Indians versus non-Indians).” Id. at 9.

The plaintiffs then explain that they served a third-party subpoena on PeopleFluent. Id. In a footnote, they state that “[d]espite repeated requests,” PeopleFluent did not respond to the subpoena for six months. Id. at 9 n.12. PeopleFluent eventually did respond, producing in two batches “a variety of analyses of Infosys’ demographic data to Plaintiffs, including analyses of Infosys’ hiring, promotion, and termination practices.” Id. at 9. (The plaintiffs assert that the first batch of information PeopleFluent provided “was

incomplete and contained over 75 corrupt and unusable files,” and that it did not provide “a more robust set of data and analyses” until a month after the first production—three weeks prior to the plaintiffs’ deadline for filing their replies in support of their partial summary judgment and class certification motions. Id. at 9 n.12.) The plaintiffs assert that Infosys did not “share these analyses with its own experts, misrepresented to Plaintiffs that these analyses did not exist, and cannot attack the analyses as a product of ‘misunderstanding’ Infosys’ data, as it does with Dr. Neumark.” Id. at 9-10.

In a footnote, the plaintiffs state:

The PeopleFluent data is simpler, considerably less burdensome, and more probative than the “barn yard” of data Infosys produced in discovery (after objecting to the burden of its production). If Infosys had produced the PeopleFluent data at the outset of discovery, Plaintiffs would, in all likelihood, have promptly moved for class certification and partial summary judgment, thereby shortening the case. Production of the data would also have fundamentally altered the focus of discovery, as Plaintiffs would have requested the search of different document custodians using different search terms than those upon which the parties ultimately agreed. Thus, Infosys’ withholding of the data has prolonged this case, irrevocably affected the focus of discovery and evidence to be presented to a jury, and has led to unnecessary expert analyses and briefing. At this juncture, there is no way to cure the prejudice to the Plaintiffs, putative class members, and the Court from Infosys electing not to disclose the data.

Id. at 10 n.13.

The plaintiffs then summarize what they say the PeopleFluent analyses demonstrated: that Infosys consistently employed “a 93% Asian workforce in the United States,” that in 2011 it promoted 14.94% of its Asian employees compared to 5.89% of non-Asian employees, in 2014 it promoted 4.97 % of its Asian workforce and only 0.11% of non-Asians, that Infosys involuntarily

terminated between 6.67% and 12.54% of is non-Asian employees and just 0.54% to 0.75% of Asians. Id. at 10-11. The plaintiffs say that PeopleFluent's analyses "use the EEO racial category 'Asian,'" which, they say, "encompasses South Asian, and the vast majority of Infosys' Asian employees are South Asian." Id. at 10 n.14. They support the last part of this assertion—that the vast majority of Infosys' Asian employees are South Asian—by citing Dr. Neumark's report. Id.

The plaintiffs next address the defendants' argument that the court must decide the motion to exclude Neumark's opinions before it decides the motion for class certification. Id. at 12. Although they agree that the court must conduct a full Daubert analysis on an expert's opinions if those opinions are critical to class certification or summary judgment, the plaintiffs state that "Plaintiffs initially relied on Dr. Neumark's testimony, after which the PeopleFluent analyses were produced." Id. at 12. The plaintiffs state, "Because the PeopleFluent analyses provide statistical support of a pattern of discrimination, the Court can decide Plaintiffs' class certification and partial summary judgment motions based on this evidence alone." Id. at 12-13.

Despite this assertion, the plaintiffs go on to argue that Neumark's methodology is relevant and reliable. Id. at 13. They cite Adams, 231 F.3d at 424 and Casteneda v. Partida, 430 U.S. 482, 496 n.17 (1977) for the proposition that Neumark used the same methodology "typically" used in employment discrimination cases and accepted by the Supreme Court and the Seventh Circuit. Id. They explain that Neumark first defined the relevant labor

market (as the U.S. computing workforce for his hiring analysis and as Infosys' own workforce for the promotion and termination analyses). Id. He then asked what the results for hiring, promotion and termination would be absent discrimination, found significant disparities and calculated the likelihood of those disparities occurring by chance. Id. From the results, he drew the inference that Infosys' practices were not race-neutral. Id. at 13-14.

The plaintiffs argue that Neumark's opinion that the disparities could not have occurred by chance is relevant "[i]n a pattern or practice case brought under *Teamsters*." Id. at 14. They assert that "the Seventh Circuit has consistently found that expert testimony regarding a statistical disparity is relevant, even if the expert's analysis did not include a multiple regression analysis accounting for factors other than race." Id. (citing Adams, 231 F.3d at 427-28; and Mister, 832 F.2d at 1431). They assert that Neumark's finding that there was a "dramatic disparity" between Infosys' hiring, promotion and termination practices as to South Asians or Indians compared to non-South Asians or Indians, combined with "other evidence," satisfies the plaintiffs' *prima facie* burden under *Teamsters*. Id. at 15.

The plaintiffs urge the court to look at what Neumark did, rather than what he or any other expert might have done. Id. They assert that what Neumark did was perform a statistical analysis showing that the disparities he observed could not have occurred by chance, and that he "was not required to perform regression analyses to exclude the hypothetical possibility that unknown non-discriminatory factors caused the disparity." Id. at 16 (citing



Adams, 231 F.3d at 427; Mister, 832 F.2d at 1431). In response to Infosys’ argument that non-discriminatory factors might account for some of the disparities, the plaintiffs assert that Neumark’s opinion is relevant to helping them establish their *prima facie* case, and assert that Infosys had a burden to produce evidence of the non-discriminatory reasons and “failed to carry its burden.” Id.

The plaintiffs then turn to the defendants’ argument that Neumark’s opinions are unreliable. Id. at 18. They assert that the defendants do not, “and cannot—argue that Dr. Neumark used flawed or unreliable methodology in determining the likelihood that random chance could account for the gross statistical disparities in Infosys’ hiring, promotions, and terminations, or in calculating standard deviations,” maintaining that Neumark’s opinions “are based on standard and reliable statistical analyses.” Id. They deny that Neumark offered any opinion on causation or discrimination, asserting that he simply “determined that the statistical likelihood of Infosys achieving it[s] rates of hiring, promotions, and terminations favoring South Asians by chance was less than 1 in 1 billion—a finding strongly consistent with discrimination.” Id. They assert that because Neumark did not offer any opinions on causation, “he does not need to live up to the scientific standards for offering such an opinion,” and return to their argument that regression analysis is not mandatory. Id. at 19.

The plaintiffs disagree that Neumark improperly defined the relevant labor market. Id. at 20. They say that the market he defined—“the market for

United States’ computing workforce”—is “the market in which the labor is conducted and is thus the appropriate market to assess.” Id. The plaintiffs say that the fact that “visa workers from India enter the United States labor market to fill positions does not broaden the geographic scope of the labor market, as Infosys argues.” Id.

The plaintiffs explain that a “potential employee overseas” has to go through the time-consuming and expensive process of applying for and receiving a work visa before entering the U.S. labor force, that there is a cap on the number of foreign workers who may fill U.S. positions and that the government uses a lottery system to award visas “against the cap.” Id. at 20-21. The plaintiffs claim that these “barriers to entry ensure that potential employees in other countries are not readily substitutable for potential employees in the United States, rendering foreign labor markets separate and distinct from a U.S. labor market.” Id. at 21. They accuse Infosys of “confront[ing]” the congressionally mandated labor market entry barriers by “investing in and creating an ‘inventory’ of visa holders who can fill positions in the U.S. on a moment’s notice.” Id. The plaintiffs argue that this “inventory” “does not render all potential employees in India—or other countries—readily substitutable for U.S. employees,” and assert that the “inventory” is “a product of fraud in which Infosys goes through the time consuming and expensive process of competing for and securing visas for individuals **before positions even exist.**” Id. at 21-22 (emphasis in original). The plaintiffs claim that the defendants’ own expert has admitted that this process is illegal. Id. at 22. The

plaintiffs contend that “[t]o suggest that Infosys’ conduct somehow broadens the world into a single labor market is to ignore the economic and regulatory barriers in place to ensure foreign workers—except those in Infosys’ inventory—cannot simply travel to the U.S. to fill a job,” and they reiterate the statement that the U.S. labor market is separate and distinct from foreign markets. Id. The plaintiffs also assert that the defendants’ arguments in this regard are foreclosed by the Seventh Circuit’s decision in Mister. Id.

As for the defendants’ argument that Neumark failed to distinguish between deputies and base employees, the plaintiffs indicate that Neumark corrected for that in his supplemental report, and they argue that his “inclusion of deputy promotions and terminations occurring in the United States is consistent with PeopleFluent’s analyses.” Id.

Turning to what they identify as Neumark’s “name recognition methodology,” the plaintiffs argue that while Neumark’s analysis may be “imperfect,” it was “reasonably accurate and Infosys substantially overstates the effect of the limited number of misclassifications.” Id. at 23. Again, the plaintiffs state that Neumark’s findings “are consistent with PeopleFluent’s analyses.” Id. They assert that Infosys “itself recognizes that it hires a proportion of South Asians within the United States that far exceeds the relevant labor market,” citing a statement by Infosys’ EEO/Diversity & Inclusion Team. Id.

The plaintiffs contend that Infosys has the burden to “identify a non-discriminatory explanation for the gross statistical disparities Dr. Neumark

demonstrates in Infosys' hiring practices," and thus dismiss the defendants' argument that Neumark failed to compare similarly situated individuals. Id. at 24. They say that Neumark "was not required to look at each individual applicant's qualifications," asserting in footnotes that at Phase II of the class certification process, the defendants "will be free to" challenge the qualifications of any individual class member. Id. at 24, nn.31, 32.

As for Neumark's promotions analysis, the plaintiffs describe the defendants' assertion that there is no evidence that a role change is the only way to receive a promotion as "false." Id. at 25. And they characterize Neumark's supplemental report as "simply ignor[ing] role changes, and look[ing] only to job levels." Id. They indicate that in his supplemental report, Neumark corrected for the fact that he had not included the minimum employment period of eighteen months contained in the class description. Id. at 26. And they say that while the defendants accuse Neumark of ignoring the pool of applicants who "competed" for promotions, there is no evidence that employees at Infosys "competed" for promotions. Id.

The plaintiffs assert that Neumark's supplemental report also corrects for errors in his termination analysis, focusing only on the termination rates among base hire employees. Id. at 27. They maintain that the supplemental report moots the defendants' criticism that Neumark improperly compared termination rates for Indians versus non-Indians by comparing the rate of termination for South Asian and non-South Asians "relative to all other base employees." Id. at 28. They assert that Neumark never claimed that downward

trends in performances of terminated individuals was evidence of discrimination, arguing that “[s]ubstantial evidence exists that Infosys manipulates CRR scores<sup>12</sup> to favor its South Asian workforce.” Id. They argue that Neumark’s table showing a decline in performance ratings prior to termination “does not itself prove discrimination (and Dr. Neumark does not claim that it does), it is consistent with Infosys’ practice of manipulating data to pretextually justify its discrimination aims.” Id. at 29 (citing Dkt. No. 88-2 at 62).

C. The Defendants’ Reply Brief (Dkt. No. 128)

The defendants reply that after weeks of asserting Dr. Neumark’s report showed “overwhelming statistical evidence” of discrimination, the plaintiffs have shifted tactics, arguing that he did not offer any opinions on discrimination or causation and thus that he need not meet the Daubert standard for offering an expert opinion. Dkt. No. 128 at 2-3. They assert that the plaintiffs now claim that no Daubert analysis is necessary because the court can rely on the PeopleFluent reports to decide the plaintiffs’ summary judgment and class certification motions. Id. at 3. The defendants conclude that the plaintiffs have conceded that Neumark’s opinions are irrelevant and unreliable. And they accuse the plaintiffs of trying to substitute their own “‘expert’ opinions about the uninformative reports of an Affirmative Action

---

<sup>12</sup> Appendix 5 to Neumark’s September 2016 report explains that Infosys conducts bi-annual performance reviews and “assigns each employee a Consolidated Relative Ranking (“CRR”) score of between 1E or 1+ to 4, with 1E and 1+ being the best and 4 being the worst.” Dkt. No. 88-2 at 58.

consultant.” Id. (They refer the court to their sur-reply in opposition to the plaintiff’s motion for class certification for their explanation of why the PeopleFluent documents are irrelevant and inadmissible. Id. at 3 n.2.)

The defendants dispute the plaintiffs’ recitation of the chronology of discovery that led to the plaintiffs’ claim that the defendant did not turn over PeopleFluent data and analysis. Id. at 3. They argue that all the data given to PeopleFluent, and more, had been given to the plaintiffs during discovery. Id. at 4. They assert that when the plaintiffs indicated that the data Infosys had produced was difficult to understand, Infosys “creat[ed] (at its own expense) charts identifying each spreadsheet by name and category, a data dictionary defining all 407 data fields, and ‘look up tables’ defining certain activity codes.” Id. They assert that defense counsel “created exemplars to illustrate how the data could be used and later met with Plaintiffs’ counsel *for hours* to explain how data could be linked, organized, and analyzed.” Id.

The defendants indicate that, in contrast, the information provided to PeopleFluent “included no data on time in job level or role, performance evaluations, or other variables that would allow meaningful statistical analysis of alleged discrimination.” Id. at 4-5. They note that PeopleFluent’s analysis did not distinguish “South Asians” from “Asians.” Id. at 5. They argue that the PeopleFluent reports “ignore all explanatory variables” and “do not (unlike even the simplistic analyses performed by Neumark) provide any benchmark or point of demographic comparison.” Id.

The defendants assert that Neumark’s “supplemental” February 2017 report “only confirms fundamental errors in Neumark’s original studies.” Id. at 5. They say that the “workforce populations study merely disaggregates the alleged unexamined disparities [Neumark] previously observed by state—a revision that proves companywide disparities do not hold across locations and employment groups.” Id. They assert that the supplement corrects some of the errors in the promotion and termination analyses, but that it doesn’t correct the most egregious ones, does not control for variables and “*concede[s] that his original analysis examined the wrong data.*” Id. And, they say, “Neumark self-approves a name-matching methodology without any scientific basis that likely distorts his results by several orders of magnitude.” Id.

As to the plaintiffs’ argument that the court can decide the class certification motion based on the PeopleFluent data, the defendants assert that these documents do not provide statistical support for a pattern of discrimination because they do not contain any statistical analysis at all; they argue that the PeopleFluent reports are neither admissible nor probative. Id. at 6. The defendants argue that the plaintiffs have the case law backward—that the Seventh Circuit has “consistently” concluded that expert opinions that do not control for variables are entitled to no weight. Id. at 6-7 (citing People Who Care v. Rockford Bd. of Educ., Sch. Dist. No. 205, 111 F.3d 528, 537-38 (7th Cir. 1997); Schultz v. Akzo Nobel Paints, LLC, 721 F.3d 425, 433 (7th Cir. 2013); Sheehan v. Daily Racing Form, Inc., 104 F.3d 940, 942 (7th Cir. 1997)). The defendants distinguish Adams and Mister. Id. at 7.

The defendants challenge the plaintiffs' reliance on Teamsters for the assertion that Infosys must identify non-discriminatory explanations for the racial disparities, asserting that Teamsters governs the burden of proof at trial, while Daubert governs the admissibility of expert opinions. Id. at 8.

The defendants assert that “[s]tripped of their elaborate (and erroneous) legal justifications, Neumark’s opinions boil down to just: 1) the obvious statement that a workforce largely composed (at any one time) of employees temporarily transferred from India does not mirror the U.S. labor market; and 2) his observations of non-random disparities in Infosys hiring, termination, and promotions that admittedly could be attributable to *anything*.” Id. at 8-9. The defendants say that the plaintiffs now have admitted that Neumark did not control for qualifications in his promotions analysis, asserting that they “purport to defend what Neumark did as an entirely conventional attempt to eliminate the role of chance in a statistical opinion.” Id. at 9. Arguing that that is not what Neumark did, the defendants assert that

[s]aying the composition of Infosys’s workforce is inconsistent with “chance” is like opining players on a basketball team are unusually tall: it states the obvious. In comparing the South Asian representation in Infosys’s nationwide workforce to the U.S. labor market, Neumark ignores the elephant-sized reason that Infosys’s workforce looks different: Deputees. At any given time, about 81% of Infosys’s workforce is composed of Deputees temporarily transferred from its operations in India (usually where they worked for the same client). In other words, Neumark has effectively opined that a workforce with large numbers of employees from a foreign labor market does not reflect the racial demographics of the United States labor market—an observation that does not even qualify as expert opinion. See Fed. R. Evid. Rule 702. And while defending that opinion, Plaintiffs nowhere acknowledge—let alone rebut—the finding that comparing Infosys’s workers hired in the United States



to applicants for those jobs yields no disparity favoring South Asian hires. See Siskin Nov. Rpt. 26-32 & Tables 5-9 (Dkt. No. 103-4).

Id. So, the defendants argue, Neumark’s opinion is irrelevant and unreliable “*even under the erroneous, burden-shifting test advocated by Plaintiffs.*” Id. at 10.

The defendants assert that excluding deputies does not “rehabilitate” Neumark’s analyses. Id. They argue that deputies “*play no part in external hiring at Infosys;*” that deputies are transferred from India, that they are not transferred into the same jobs that are available to “external” candidates and that they did not receive any job sought by representative plaintiffs Koehler and Parker. Id. They state that, “[a]t best, Neumark’s opinion could relate to a claim that Deputies make external hiring unnecessary; but Title VII does not obligate Infosys to *create* jobs.” Id. at 10-11.

The defendants assert that the plaintiffs “cannot restore Neumark to relevance with bald accusations that Infosys uses Deputies ‘to further its discriminatory objective of filling U.S. positions with its preferred race of employees.’” Id. at 11. The defendants say that the plaintiffs attempt to equate “(unsupported) allegations that Infosys has exceeded its fair share of H1-B visas with discriminatory motives,” and the defendants argue that these are different concepts. Id. They maintain that visa limits protect the domestic labor force from competition from foreign workers entering the U.S. market—arguably a practice non-discriminatory to U.S. workers—and say that the plaintiffs’ “assumption that race is the *only* reason Infosys could want to

temporarily staff U.S. projects with existing, skilled, trained and high-performing employees from India is, itself, a race-based assumption.” Id. at 11-12.

The defendants say the plaintiffs are left with arguing only that Neumark “provides a significant step in the proof when he opines chance cannot explain aggregate racial disparities in hiring, promotions, and terminations at Infosys.” Id. at 12. The defendants argue that even this pared-down defense of Neumark’s work does not make it relevant, reiterating that the name-matching methodology is unreliable and the failure to control for other variables is fatal. Id. They argue that the unreliability of the name-matching methodology is material, with the potential to misidentify hundreds of names. Id. at 13. They assert that Neumark’s own opinion that he thinks his method works “rather well” is “say-so,” not science. Id. They reiterate that failure to control for non-discriminatory variables leaves the plaintiffs with nothing but a statistical disparity that could not have occurred by chance. Id. at 13-14. They assert that the plaintiffs did not address their argument that Neumark failed to compare applicants for the same jobs (and thus failed to compare similarly situated individuals). Id. at 14.

As for promotions and terminations, the defendants assert that Neumark’s adjustment to his report amounts to a concession that Neumark based the original analyses on the wrong data, and that “[d]isparities that supposedly showed company-wide discrimination falsely assumed Infosys almost never terminates Deputies.” Id. at 14-15. They reiterate that Neumark

failed to control for any variable other than race. Id. at 15-16. And they assert that by breaking the termination data down by state, Neumark showed that his earlier analyses aggregating the data was flawed, because the state-by-state data showed little meaningful disparity in termination for one-third of the states he examined, while the aggregate showed much higher standard deviation. Id. at 16.

#### **V. The Timing of Deciding the *Daubert* Motion**

At the hearing on the motion to exclude Neumark's testimony, the court told the parties that "putting the expert testimony motion on for a hearing today is step one in being able to address the class certification motion and summary judgment motions." Dkt. No. 204 at 5, lines 16-18. The court first stated that under two Seventh Circuit cases—Am. Honda, 600 F.3d at 815-16 and Messner v. N. Shore Univ. HealthSystem, 669 F.3d 802, 812 (7th Cir. 2012)—it was required to decide the motion to exclude Neumark's testimony before deciding the class certification motion if his testimony was critical to that motion. Id. at 13, lines 12-23. It observed that the plaintiffs had argued Neumark's testimony was not critical because they were relying on different evidence. Id. at 14, lines 10-15. The court found, however, that "the plaintiffs have, in fact, in numerous places in the class certification motion referred to Dr. Neumark's opinion." Id. at lines 21-23. The court concluded:

So the idea that well, you know, we're not relying on him for class certification anymore, Dr. Neumark being him, is belied by the pleadings themselves. Dr. Neumark is all over the class certification proceedings. And so simply on that basis, I think that it is critical—Dr. Neumark's testimony is critical to the class certification motion or even at a lower level, there is certainly some question as to

whether it's critical. And under *Messner*, that means I should decide it.

Id. at 15, lines 9-16. The following puts flesh on that bony conclusion.

“Before deciding whether to allow a case to proceed as a class action . . . a judge should make whatever factual and legal inquiries are necessary under Rule 23.” Szabo v. Bridgeport Machs., Inc., 249 F.3d 672, 676 (7th Cir. 2001). “And if some of the considerations under Rule 23(b)(3) . . . overlap the merits . . . then the judge must make a preliminary inquiry into the merits.” Id. Plaintiffs cannot obtain class certification “just by hiring a competent expert,” and a district judge “may not duck hard questions by observing that each side has some support, or that considerations relevant to class certification also may affect the decision on the merits.” West v. Prudential Sec., Inc., 282 F.3d 935, 938 (7th Cir. 2002). “Tough questions must be faced and squarely decided, if necessary by holding evidentiary hearings and choosing between competing perspectives.” Id.

Consequently, the Seventh Circuit has held that “when an expert’s report or testimony is critical to class certification . . . a district court must conclusively rule on any challenge to the expert’s qualifications or submissions prior to ruling on a class certification motion. That is, the district court must perform a full *Daubert* analysis before certifying the class if the situation warrants.” Am. Honda 600 F.3d at 815-16. “The court must . . . resolve any challenge to the reliability of information provided by an expert if that information is relevant to establishing any of the Rule 23 requirements for class certification.” Id. at 816. See also Howard v. Cook Cty. Sheriff’s Office,

989 F.3d 587, 601 (7th Cir. 2021) (finding district court's reliance on excluded expert's opinions in analyzing commonality requirement of Rule 23 "improper" and citing Am. Honda). "If a district court has doubts about whether an expert's opinions may be critical for a class certification decision, the court should make an explicit *Daubert* ruling." Messner, 669 F.3d at 812. "An erroneous *Daubert* ruling excluding non-critical expert testimony would result, at worst, in the exclusion of expert testimony that did not matter. Failure to conduct such an analysis when necessary, however, would mean that the unreliable testimony remains in the record, a result that could easily lead to reversal on appeal." Id.

Rule 23(a) allows a member of a class to sue as a representative of that class only if:

- (1) the class is so numerous that joinder of all members is impracticable;
- (2) there are questions of law or fact common to the class;
- (3) the claims or defenses of the representative parties are typical of the claims or defenses of the class; and
- (4) the representative parties will fairly and adequately protect the interests of the class.

In addition, Rule 23(b) requires that for a case to proceed as a class action, one of three circumstances must exist; one of which is where "the court finds that the questions of law or fact common to class members predominate over any questions affecting only individual members, and that a class action is superior to other methods for fairly and efficiently adjudicating the controversy." Fed. R. Civ. P. 23(b)(3).

The plaintiffs' class certification motion states that they seek class

certification for the determination of “whether Infosys engaged in a pattern or practice of discrimination; and the availability of punitive damages and injunctive relief.” Dkt. No. 88 at 7-8. It states that to prove the pattern or practice of discrimination, the plaintiffs will “rely partly on statistical evidence of discrimination, which is common to the class.” Id. The plaintiffs cite Neumark’s report twenty-six times in the opening brief in support of the class certification motion. Specifically, they cite his report in support of their assertion that they meet the Rule 23(a)(1) numerosity requirement, id. at 21-22; and in support of their Rule 23(b)(3) argument that common questions predominate over individual issues, id. at 25-26 (citing analysis in Section II of the facts common to the class, which repeatedly cites from Neumark’s September 2016 report).

To be allowed to represent the class, the plaintiffs must prove numerosity (among other things). For the case to proceed as a class action, the plaintiffs must prove that common questions predominate over individual ones. In their opening brief, the plaintiffs relied heavily on Neumark’s opinions to argue these Rule 23 prerequisites. Neumark’s opinion was relevant—arguably critical—to their ability to prove those prerequisites.

The defendants argued in their opposition to the motion to certify the class that the court should deny the class certification motion because the plaintiffs could not demonstrate under Rule 23(a) that there are common questions of law or fact. Dkt. No. 103 at 9. Citing Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338 (2011), the defendants argue that the plaintiffs cannot

provide convincing proof of a pattern or practice of discrimination and assert that Neumark's statistical evidence is the kind of evidence the Dukes Court rejected as insufficient for that purpose. Id. at 10-12. Whether the defendants are correct that Dukes mandates denial of the class certification motion, they are correct that in a case where the plaintiffs have alleged a pattern or practice of discrimination, the commonality question overlaps with the merits of the claims. Dukes, 564 U.S. at 352. They also are correct that in concluding that the plaintiffs in Dukes had not proved commonality, the Supreme Court found lacking regression analyses that showed significant statistical disparities between Wal-Mart's workforce and the allegedly relevant workforce and concluded that they could be explained only by gender discrimination. Id. at 356. Under Dukes, the plaintiffs in this pattern-or-practice suit—who are alleging that Infosys had a pattern or practice of discriminating in thousands of hiring, promotion and firing decisions—must present evidence of “some glue holding the alleged *reasons* for all those decisions together,” or “it will be impossible to say that examination of all the class members' claims for relief will produce a common answer to the crucial question *why was I disfavored.*” Id. at 352. Neumark's analysis is part of the evidence the plaintiffs cite, so it is part of the evidence the court would be required to consider in ruling on the Rule 23(a)(2) commonality factor.

The plaintiffs opposed the motion to exclude Neumark's opinions by arguing that PeopleFluent did not timely respond to their third-party subpoena; they argue that the defendants' failure to produce in discovery the information

the plaintiffs subpoenaed and obtained from PeopleFluent caused the plaintiffs to delay in filing the class certification motion and their motion for partial summary judgment. This argument has nothing to do with whether the court must conduct a Daubert analysis of Neumark's opinions before deciding the class certification motion. If PeopleFluent did not timely respond to the plaintiffs' third-party subpoena, they could have raised that with Judge Jones (and perhaps they did). They could have filed a motion to compel. They could have sought an extension of the deadline for filing the motion to certify the class (although they explained at the motion hearing that they did not want to "slow down a decision on class certification"). Dkt. No. 204 at 51, lines 23-24.

The plaintiffs have not stated that they would not have retained Neumark, or that they would not have relied on his opinions in the class certification motion, if they'd had the PeopleFluent information. On the contrary—at the motion hearing, the plaintiffs said that by discussing the PeopleFluent data in the reply brief, "we did not mean to suggest that we should withdraw Dr. Neumark's analysis and substitute it with PeopleFluent's analysis," dkt. no. 204 at 44, lines 11-14; they asserted that "of course" they thought Neumark's analyses were "relevant in order for class certification," dkt. no. 204 at 51, lines 20-21.

The plaintiffs' assertion in the reply brief that the court need not conduct a Daubert analysis of Neumark's opinions because it can decide the class certification and partial summary judgment motions based on the PeopleFluent data also misses the mark. As the court explained at the hearing and again in



this order, the class certification motion relies on Neumark's opinion, not the PeopleFluent data. (The plaintiffs' motion for partial summary judgment also relies heavily on Neumark's September 2016 opinions. Dkt. No. 86.)

The court must conduct a Daubert analysis of Neumark's work before it rules on the class certification motion (and on the plaintiffs' motion for partial summary judgment).

## **VI. Daubert Analysis**

### **A. Applicable Law**

Federal Rule of Evidence 702 and Daubert govern the admissibility of expert testimony. Rule 702 says that a witness "may" testify as an expert if the witness is "an expert by knowledge, skill, experience, training, or education;" the expert's knowledge "will help the trier of fact to understand the evidence or to determine a fact in issue;" the expert's testimony is "based on sufficient facts or data" and is "the product of reliable principles and methods;" and the expert "has reliably applied the principles and methods to the facts of the case." This rule "assign[s] to the trial judge the task of ensuring that an expert's testimony both rests on a reliable foundation and is relevant to the task at hand."

Daubert, 509 U.S. at 597. "*Daubert* and Rule 702 apply 'to social science experts,' just as they apply 'to experts in the hard sciences.'" Howard, 989 F.3d at 601 (quoting Tyus v. Urban Search Mgmt., 102 F.3d 256, 263 (7th Cir. 1996)).

Under Rule 702 and Daubert, a court must engage in a three-step inquiry before it admits expert witness testimony. Gopalratnam v. Hewlett-Packard Co., 877 F.3d 771, 779 (7th Cir. 2017).

In performing its gatekeeper role under Rule 702 and *Daubert*, “the district court must engage in a three-step analysis before admitting expert testimony. It must determine whether the witness is qualified; whether the expert’s methodology is scientifically reliable; and whether the testimony will ‘assist the trier of fact to understand the evidence or to determine a fact in issue.’” In other words, the district court must evaluate: (1) the proffered expert’s *qualifications*; (2) the *reliability* of the expert’s methodology; and (3) the *relevance* of the expert’s testimony.

Id. (emphasis in original) (internal citations omitted).

The party seeking to introduce expert witness testimony bears the burden of showing by a preponderance of the evidence that the witness’ testimony satisfies the Daubert standard. Id. at 782.

1. *Qualifications*

Rule 702 states that an expert witness may be qualified as an expert “by knowledge, skill, experience, training or education.” The fact that an expert may offer opinions that are not based on firsthand knowledge or observation “is premised on an assumption that the expert’s opinion will have a reliable basis in the knowledge and experience of his discipline.” Daubert, 509 U.S. at 592. An expert need not have particular academic credentials to be qualified; “anyone with relevant expertise enabling him to offer responsible opinion testimony helpful to judge or jury may qualify as an expert witness.” Tuf Racing Prods, Inc. v. Am. Suzuki Motor Corp., 223 F.3d 585, 591 (7th Cir. 2000). “The question [the court] must ask is not whether an expert witness is qualified in

general, but whether his ‘qualifications provide a foundation for [him] to answer a specific question.’” Gayton v. McCoy, 593 F.3d 610, 617 (7th Cir. 2010) (quoting Berry v. City of Detroit, 25 F.3d 1342, 1351 (6th Cir. 1994)). The court must look at each of the expert’s conclusions individually “to see if he has the adequate education, skill, and training to reach them.” Id. “[A] court should consider a proposed expert’s full range of practical experience as well as academic or technical training when determining whether that expert is qualified to render an opinion in a given area.” Smith v. Ford Motor Co., 215 F.3d 713, 718 (7th Cir. 2000).

## 2. *Reliability*

In Gopalratnam, the Seventh Circuit discussed in depth the reliability requirement:

According to our circuit’s precedent, courts should evaluate the reliability of a qualified expert’s testimony by considering, amongst other factors: “(1) whether the proffered theory can be and has been tested; (2) whether the theory has been subjected to peer review; (3) whether the theory has been evaluated in light of potential rates of error; and (4) whether the theory has been accepted in the relevant scientific community.” Krik [v. Exxon Mobil Corp.], 870 F.3d [669,]at 674 [(7th Cir. 2017)] (quoting Baugh v. Cuprum S.A. de C.V., 845 F.3d 838, 844 (7th Cir. 2017)). In addition, the Rule 702 advisory committee’s note to the 2000 amendment outlines other benchmarks relevant in assessing an expert's reliability:

(5) whether “maintenance standards and controls” exist; (6) whether the testimony relates to “matters growing naturally and directly out of research they have conducted independent of the litigation,” or developed “expressly for purposes of testifying”; (7) “[w]hether the expert has unjustifiably extrapolated from an accepted premise to an unfounded conclusion”; (8) “[w]hether the expert has adequately accounted for obvious alternative explanations”; (9) “[w]hether the expert is being as careful as he would be in his regular professional work

outside his paid litigation consulting”; and (10) “[w]hether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give.”

*Fuesting v. Zimmer, Inc.*, 421 F.3d 528, 534–35 (7th Cir. 2005), *opinion vacated in part on reh’g*, 448 F.3d 936 (7th Cir. 2006) (quoting Fed. R. Evid. 702 advisory committee’s note to 2000 amendment).

“Importantly, this list is neither exhaustive nor mandatory.” [*C.W. ex rel. Wood v. Textron*, 807 F.3d [827,] at 835 [(7th Cir. 2015)]; see also *Kumho Tire [v. Carmichael]*, 526 U.S. [137,] at 150, 119 S.Ct. 1167 [(1999)] (“*Daubert* makes clear that the factors it mentions do not constitute a ‘definitive checklist or test.’” (quoting *Daubert*, 509 U.S. at 593, 113 S. Ct. 2786)); *Krik*, 870 F.3d at 674 (“Despite the list, we have repeatedly emphasized that ‘no single factor is either required in the analysis or dispositive as to its outcome.’” (quoting *Smith v. Ford Motor Co.*, 215 F.3d 713, 719 (7th Cir. 2000))); *United States v. Cruz-Velasco*, 224 F.3d 654, 660 (7th Cir. 2000) (“Although the *Daubert* Court identified a number of factors to be considered when evaluating the admissibility of expert testimony ... these factors do not establish a definitive checklist.”). Instead, “a trial court *may* consider one or more of the more specific factors that *Daubert* mentioned when doing so will help determine that testimony’s reliability.” *Kumho Tire*, 526 U.S. at 141, 119 S. Ct. 1167.

Ultimately, “there are many different kinds of experts, and many different kinds of expertise.” *Id.* at 150, 119 S. Ct. 1167. The test of reliability, therefore, “is ‘flexible,’ and *Daubert*’s list of specific factors neither necessarily nor exclusively applies to all experts or in every case.” *Id.* at 141, 119 S. Ct. 1167 (quoting *Daubert*, 509 U.S. at 594, 113 S. Ct. 2786); see also *Textron*, 807 F.3d at 835 (“Ultimately, reliability is determined on a case-by-case basis.”). Rather, “[t]he district court may apply these factors flexibly as the case requires.” *Krik*, 870 F.3d at 674; see also *Kumho Tire*, 526 U.S. at 142, 119 S. Ct. 1167 (“[T]he law grants a district court the same broad latitude when it decides *how* to determine reliability as it enjoys in respect to its ultimate reliability determination.”). In the end, “the gatekeeping inquiry must be ‘tied to the facts’ of a particular ‘case,’” *Kumho Tire*, 526 U.S. at 150, 119 S. Ct. 1167 (quoting *Daubert*, 509 U.S. at 591, 113 S. Ct. 2786), and “the reliability analysis should be geared toward the precise sort of testimony at issue and not any fixed evaluative factors.” *Lees [v.*

*Carthage College*], 714 F.3d [516,] at 521 [(7th Cir. 2013)].

At the same time, this flexibility is not without limit. “[T]he district court’s role as gatekeeper does not render the district court the trier of all facts relating to expert testimony. ... The jury must still be allowed to play its essential role as the arbiter of the weight and credibility of expert testimony.” *Stollings v. Ryobi Techs., Inc.*, 725 F.3d 753, 765 (7th Cir. 2013) (citations omitted). Rather, “Rule 702’s reliability elements require the district judge to determine only that the expert is providing testimony that is based on a correct application of a reliable *methodology* and that the expert considered sufficient data to employ the methodology.” *Id.* at 766 (emphasis added). This examination “does not ordinarily extend to the reliability of the *conclusions* those methods produce—that is, whether the conclusions are unimpeachable.” *Id.* at 765 (emphasis added). In other words, “[a]n expert may provide expert testimony based on a valid and properly applied methodology and still offer a conclusion that is subject to doubt. It is the role of the jury to weigh these sources of doubt.” *Id.* at 766.

The focus, therefore, “must be solely on principles and methodology, not on the conclusions that they generate.” *Daubert*, 509 U.S. at 595, 113 S. Ct. 2786; *see also Ford Motor Co.*, 215 F.3d at 718 (“[W]e emphasize that the court’s gatekeeping function focuses on an examination of the expert’s methodology.”). “The soundness of the factual underpinnings of the expert’s analysis and the correctness of the expert’s conclusions based on that analysis are factual matters to be determined by the trier of fact, or where appropriate, on summary judgment.” *Ford Motor Co.*, 215 F.3d at 718; *see also Manpower, Inc. v. Ins. Co. of Pa.*, 732 F.3d 796, 806 (7th Cir. 2013) (“Reliability ... is primarily a question of the validity of the methodology employed by an expert, not the quality of the data used in applying the methodology or the conclusions produced.”). “The district court usurps the role of the jury, and therefore abuses its discretion, if it unduly scrutinizes the quality of the expert’s data and conclusions rather than the reliability of the methodology the expert employed.” *Manpower*, 732 F.3d at 806.

“This is not to say that an expert may rely on data that has no quantitative or qualitative connection to the methodology employed.” *Id.* at 808. Indeed, Rule 702 explicitly requires that expert testimony be “based on sufficient facts or data.” Fed. R. Evid. 702. In the “quantitative” sense, “‘sufficient facts or data’ means ‘that the expert considered sufficient data to employ the methodology’”; “an opinion about an average gross sales price,” for

example, “could not be reliably supported by evidence relating to sales to only one customer ‘because a single observation does not provide a sufficient basis for calculating an average.’” *Manpower*, 732 F.3d at 808 (quoting *Stollings*, 725 F.3d at 766). To be “qualitatively” adequate, “an expert must employ ‘those kinds of facts or data’ on which experts in the field would reasonably rely.” *Id.* at 809 (quoting Fed. R. Evid. 703).

We have recognized that the line between conclusions and methodology “is not always an easy line to draw.” *Id.* at 806. “[C]onclusions and methodology are not entirely distinct from one another. Trained experts commonly extrapolate from existing data.” *Gen. Elec. Co. v. Joiner*, 522 U.S. 136, 146, 118 S. Ct. 512, 139 L.Ed.2d 508 (1997). Nevertheless, “[t]he critical inquiry is whether there is a *connection* between the data employed and the opinion offered; it is the opinion connected to existing data ‘only by the *ipse dixit* of the expert’ that is properly excluded under Rule 702.” *Manpower*, 732 F.3d at 806 (quoting *Joiner*, 522 U.S. at 146, 118 S.Ct. 512) (first emphasis added). Said another way, there must be a “rational connection between the data and the opinion.” *Id.* at 809.

Gopalratnam, 877 F.3d at 779–81.

### 3. *Relevance*

Relevant evidence is “that which has ‘any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.’” Daubert, 509 U.S. at 587 (quoting Fed. R. Evid. 401). Rule 702’s “basic standard of relevance thus is a liberal one.” Id. The Rule 702 requirement that the expert’s evidence or testimony must “assist the trier of fact to understand the evidence or determine a fact in issue” “goes primarily to relevance.” Id. at 591.

B. Analysis

1. *Dr. Neumark's Qualifications*

As the court noted at the hearing, the defendants have not challenged Neumark's general qualifications. Dkt. No. 204 at 20, lines 5-13; 23, lines 5-7. They do not dispute that he is a professor of economics at UC-Irvine or his assertion that he has done "extensive research on labor market discrimination, including methods for measuring and testing for discrimination that have been adopted by many other researchers." Dkt. No. 88-2 at 4. They do not dispute his prior positions on the Federal Reserve Board, the University of Pennsylvania, Michigan State University or the Public Policy Institute of California. Id. They do not dispute his extensive publications (the listing of which consumes nine pages of his report). Id. at 66-75. In his extensive deposition, the defendants did not enquire into Neumark's qualifications as a labor economist, a scholar or a researcher. Dkt. No. 97-1.

Rather, the defendants argue that Neumark is not qualified to identify the race or ethnicity of an employee or applicant based on name. Dkt. No. 204 at 23, lines 8-9 ("we do dispute that he is qualified to perform a name matching analysis"). At Neumark's deposition, the following exchange occurred:

Q.: Have you ever used name-matching in any other case?

A.: No.

Q.: Do you have any expertise in what may sound like an Indian name?

A.: No. That's why I—that's why I did what I did and didn't do what you're suggesting here. Because I didn't have to rely on what sounds like an Indian name.

Q.: Well, you did it with respect to other kinds of names. I mean, you don't—you don't have any expertise in name recognition at all, correct?



A.: I don't have any expertise in it, no.

Dkt. No. 97-1 at 38, Tr. Page 143, lines 7-20.

When asked whether it would be significant if fourteen percent of the Kenexa applicants he identified as Indian self-identified as non-Asian, Neumark replied:

Yeah. I mean, these are—these are different ways you can define ethnicity as we discussed. I'm doing it based on names. It's not perfect. I don't know—I have not studied the self-identification data and how often it's reported. I have no—you know, if I was—if you guys asked about Chinese calling themselves non-Asian, I'd probably be a little more—that would seem weirder. I just don't know. I'm in complete ignorance here about if Indians self-identify as—as Asian ethnicity. I—I just don't know.

Dkt. No. 97-1 at 37, Tr. Page 140, lines 1-10.

This lack of expertise is a problem under Daubert.

Neumark's comparisons of Infosys' employee and applicant demographic compositions with what he defined as the relevant labor market depended on an accurate representation of Infosys' employee demographic composition and an accurate representation of the demographic composition of Infosys' applicant pool. Neumark did not have data sufficient to allow him to accurately determine those demographic compositions.

Neumark compared data from the ACS, which provides national origins classifications and thus allows people to identify as South Asian or Indian, with data from Infosys, which his report said<sup>13</sup> provided only the broad classification of "Asian," and did not "differentiate between someone of Indian or Southeast

---

<sup>13</sup> The defendants assert that for applicants managed in the Kenexa software, there was self-identified race information. Dkt. No. 98 at 12.



Asian descent and someone of, for example, Chinese descent; both groups would be coded as Asian.” Dkt. No. 88-2 at 43. Although the Infosys data identified Infosys employees who reported being born in India or in countries Neumark considered to be South Asian, it did not identify employees of Indian or South Asian descent not born in India or in a country Neumark considered South Asian, and it did not identify some applicants of Indian or South Asian descent. Id.

Because Neumark did not have the data that would have allowed him to accurately determine the Infosys employee and applicant demographic compositions, he created a methodology for identifying people of Indian and South Asian descent: he looked at the last names of Infosys employees who reported being born in India or one of the countries he considered to be South Asian<sup>14</sup> and “use[d] these names to identify South Asians or Indians among records on employees not born in these countries, or among records on applicants.” Id. He then made a list of last names associated with at least two Infosys employees, where at least one employee was born in India or a country he identified as South Asian and at least one employee was not born in a South

---

<sup>14</sup>Neumark also based his conclusions about the percentages of South Asian employees or applicants on the assumption that “South Asians include the following ethnicities: Indian, Sri Lankan, Bangladeshi, Pakistani, Burmese, Nepalese, and Afghan.” Dkt. No. 88-2 at 13 n.8. The second amended complaint does not define “South Asian,” although it initially defined the proposed class as individuals who were not “of South Asian race or Indian, Bangladeshi, or Nepalese origin” who were not hired, not promoted or fired. Dkt. No. 19-2 at ¶121. Neumark did not explain why he included other countries—Sri Lanka, Pakistan, Burma or Afghanistan—in his definition of “South Asian,” or how he determined that the countries he listed were “South Asian” and that others were not.

Asian country or India. Id. at 44. From *that* list, he “identified surnames that [did] not appear to be Indian.” Id. And he quantified how many of the employees with those surnames were born in India or countries he considered to be South Asian. Id.

Despite his education, experience and expertise in labor economics, Neumark has no expertise in name recognition. There is no evidence that he was qualified to create a method for determining whether someone is or is not Indian or South Asian based on surname. Yet he not only did so, but he used that method to determine the percentage of Infosys’ United States-based workforce that purportedly consisted of South Asians, *dkt. no. 88-2 at 11*; the percentage of Infosys’ hires in the United States that purportedly were South Asian, id. at 11-12; the percentage of Infosys’ hires in the United States that purportedly were Indian, id. at 12-13; the percentage of Infosys’ United States-based workforce that purportedly consisted of Indians, id. at 14-15; the percentages of promotions that purportedly went to Indian employees, id. at 29; and the percentage of terminations of purportedly Indian/South Asian employees, id. at 31. Underlying each of Neumark’s comparisons is his conclusion about how many employees or applicants were of Indian national origin or South Asian race/ethnicity, and he drew that conclusion using name recognition—an endeavor for which he admittedly had no expertise.

The plaintiffs did not address this aspect of Neumark’s qualifications—or, more accurately, this aspect of his lack of qualifications. Instead, they asserted that there “is no perfect way to identify South Asians from the data,” cited

Neumark's report for his explanation of what he had done, and asserted that his analysis, "though imperfect, is reasonably accurate." Dkt. No. 123 at 23. That does not address Neumark's lack of qualifications; those assertions go to the reliability of his methodology. The plaintiffs have presented no evidence that Neumark was qualified to identify an employee's or applicant's race by name.

The plaintiffs' reply argument that because Neumark did not opine on causation his analyses need not meet the requirements of Daubert is a non-starter. Neumark's report says that Infosys "Disproportionately Hired South Asians and Indians," dkt. no. 88-2 at 7, that it disproportionately promoted South Asians and Indians, id. at 26, and that it disproportionately terminated non-South Asians and non-Indians, id. at 30. These are conclusions—opinions. In the motions for class certification and partial summary judgment, the plaintiffs refer repeatedly to Neumark's "expert" report. Neumark himself says the plaintiffs hired him "as a statistical expert" to evaluate whether the data were consistent with discrimination. Dkt. No. 88-2 at 4-5. The plaintiffs retained and presented Neumark as an expert and, until the defendants criticized his opinions, relied on him as an expert. His opinions and methods must pass muster under Rule 702 and Daubert, and he must be qualified as an expert under Rule 702 and Daubert. In the area of identifying an individual's race by name, his method does not pass muster and he is not qualified as an expert.

Arguably, this conclusion could end the analysis. But the plaintiffs claim that “[m]any of Infosys’ criticisms center on ancillary issues that relate to a subset of Dr. Neumark’s analyses and that stem from the complexity of the data Infosys produced in discovery.” Dkt. No. 123 at 8. The court does not know whether the plaintiffs consider Neumark’s race identification methodology to be “ancillary” (although the court does not). And the plaintiffs argued at the motion hearing that “to exclude our expert and allow [the defendants] to go forward without our expert performing statistical disparities showing the statistical disparities to the jury would not be right. It’s not fair.” Dkt. No. 204 at 45, lines 13-16. In an abundance of caution, the court will address the remaining Daubert factors—reliability and relevance.

## 2. *Reliability*

In 2000, Fed. R. Evid. 702 was amended in the wake of Daubert and the cases that had applied it. The Committee Notes that accompanied that amendment reiterate the Daubert Court’s caution that its list of factors for trial courts to use in assessing reliability—whether the expert’s theory or methodology has been tested, whether the theory or methodology has been peer reviewed, whether the theory or methodology has been evaluated for error rates and whether the theory or methodology has been accepted in the relevant scientific community—was not exhaustive. The Committee Notes explained that before and after Daubert, courts had found factors other than those listed in Daubert “relevant in determining whether expert testimony is sufficiently

reliable to be considered by the trier of fact.” Among these “other factors,” the Rules Committee listed:

(3) Whether the expert has adequately accounted for obvious alternative explanations. *See Claar v. Burlington N.R.R.*, 29 F.3d 499 (9th Cir. 1994) (testimony excluded where the expert failed to consider other obvious causes for the plaintiff’s condition). *Compare Ambrosini v. Labarraque*, 101 F.3d 129 (D.C. Cir. 1996) (the possibility of some uneliminated causes presents a question of weight, so long as the most obvious causes have been considered and reasonably ruled out by the expert).

and

(4) Whether the expert “is being as careful as he would be in his regular professional work outside his paid litigation consulting.” *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997). *See Kumho Tire Co. v. Carmichael*, 119 S. Ct. 1167, 1176 (1999) (*Daubert* requires the trial court to assure itself that the expert “employs in the courtroom the same level of intellectual rigor that characterizes the practice of an expert in the relevant field”).

The defendants argue that Neumark’s opinions are unreliable because he did not control for non-discriminatory variables, *dk.* no. 98 at 23-25; because he did not use the proper labor market for comparison, *id.* 25-29; because his race definition method is “irrational, arbitrary, and unscientific,” *id.* at 29-32; because his analysis is based on faulty assumptions, *id.* at 32-33; because his promotion analysis does not consider similarly situated employees, *id.* at 34-35; and because his termination analysis lacks evidentiary support, *id.* at 35-37.

In determining whether the expert’s opinion or testimony is reliable, “[t]he focus . . . must be solely on principles and methodology, not on the conclusions that they generate.” *Daubert*, 509 U.S. at 595. “[T]he key to the

gate [over which the district court plays the role of gatekeeper] is not the ultimate correctness of the expert's conclusions,' but rather 'the soundness and care with which the expert arrived at [his] opinion.'" Burton v. E.I. du Pont de Nemours & Co., Inc., 994 F.3d 791, 826 (7th Cir. 2021) (quoting Schultz, 721 F.3d at 431). "So long as the principles and methodology reflect reliable scientific practice, "[v]igorous cross-examination, presentation of contrary evidence, and careful instruction on the burden of proof are the traditional and appropriate means of attacking shaky but admissible evidence.'" Id. (quoting Schultz, 721 F.3d at 431, which quoted Daubert, 509 U.S. at 596).

The Seventh Circuit has acknowledged that it can be hard to draw the line between conclusions and methodology. Gopalratnam, 877 F.3d at 781 (quoting Manpower, Inc., 732 F.3d at 806). It is difficult to draw that line here, but it seems to the court that all but two of the defendants' criticisms go to the weight that a factfinder might give Neumark's conclusions at trial, rather than challenges to the reliability of his methodology. The defendants' criticisms that Neumark incorrectly defined the relevant labor market for comparison purposes,<sup>15</sup> that he based his analyses on faulty assumptions and that his

---

<sup>15</sup> In Hazelwood, the Supreme Court found that the district court had erred in its statistical analysis of the relevant labor market. Hazelwood, 433 U.S. at 308. In Chi. Miniature, the Seventh Circuit said that a "central task for the district court was to define the relevant labor market." Chi. Miniature, 947 F.2d at 295. These cases discussed the trial court's determinations regarding the relevant labor market in the context of trial, not in the context of analyzing expert opinions and testimony under Daubert. The Seventh Circuit has held that the determination of the proper labor market for comparison purposes is "a question of fact for the trial court to consider." Adams, 231 F.3d at 423 (citing Hazelwood, 433 U.S. at 310-12).

promotion and termination analyses were flawed are topics that, if the case were to go to trial and Neumark were to testify, might be fertile ground for testing through the “vigorous cross examination” and “presentation of contrary evidence” (such as the testimony of the defendants’ own expert) that the Daubert Court characterized as appropriate, but not criticisms that warrant exclusion of his opinions. That leaves Neumark’s failure to control for variables that may have demonstrated non-discriminatory bases for disparities and his race definition methodology.

a. Failure to Consider Variables Other Than Race

Neumark admitted at his deposition that he did not consider any variables other than race. He admitted that he had not considered the role of performance ratings in his promotion and termination analyses, indicating that he’d been asked only to “look at the patterns in employment hiring, termination, and promotions.” Dkt. No. 97-1 at 12, Tr. Page 39, line 25-Tr. Page 40, lines 1-11. He testified that he had not “looked at kind of the characteristics of people, deputies or otherwise, that might be influenced in the outcomes. I’ve simply look[ed] at the descriptive evidence on the outcomes and what the patterns looked like.” Id. at 15, Tr. Page 51 at lines 12-16. He testified that he had not “gone into the institutional details of how [the defendants]—how they do their recruiting and hiring in much detail at all,” testifying that he simply “focused on the data, who works there, who’s been hired, and what the statistical patterns are in those data.” Id. at 17, Tr. Page 58 at lines 12-16. He testified that “for the analysis that’s in [the September 2016] report, none of

that—none of the figuring out what qualifications matter and incorporating it into the analysis was done.” Id. at 26, Tr. Page 96 at lines 11-14. He testified that he did not “look at the ethnic identification information from the applicant data.” Id. at 38, Tr. Page 141 at lines 16-18. Neumark testified that he did not conduct “an analysis of factors that could explain the difference in promotion rates. I was documenting the differences in promotion rates.” Id. at 43-44, Tr. Page 164 at lines 24-25 to Tr. Page 165 at lines 1-2. He did not look at the factors determining promotion. Id. at 48, Tr. Page 181 at lines 20-21. He testified that he was asked “to look at the patterns of the outcomes by Indian, non-Indian, South Asian, non-South Asian and restricted to that.” Id. at 62, Tr. Page 239 at lines 2-4.

The defendants attack Neumark’s failure to consider variables other than race in a couple of ways. First, they argue that by failing to consider variables other than race, Neumark acted contrary to principles he’d approved in his scholarly writing and about which he’d testified in at least one other case. Arguably, like the defendants’ criticisms of Neumark’s definition of the relevant labor market, their assertion that he relied on faulty assumptions, and their claims that his promotion and termination analyses were faulty, this argument could go to the weight a factfinder ought to give Neumark’s conclusions. But the Seventh Circuit has twice concluded that statistical evidence was inadmissible under Daubert because the expert had not used the same care in his litigation work as he used in his non-litigation professional work.



In Sheehan, 104 F.3d 940, authored by Judge Posner with Judges Eschbach and Evans rounding out the panel, Judge Posner cited Daubert in offering the following criticism of the statistician who authored the affidavit the plaintiff presented to show a *prima facie* case of age discrimination:

The expert's failure to make any adjustment for variables bearing on the decision whether to discharge or retain a person on the list other than age—his equating a simple statistical correlation to a causal relation (“of course, if age had no role in termination, we should expect that equal proportions of older and younger employees would be terminated”—true only if no other factor relevant to termination is correlated with age)—indicates a failure to exercise the degree of care that a statistician would use in his scientific work, outside of the context of litigation.

Sheehan, 104 F.3d at 942.

People Who Care, 111 F.3d 528, is another Judge Posner-authored decision, issued only three months after Sheehan; Judges Bauer and Kanne made up the panel. Judge Posner opined that an achievement gap study prepared for a desegregation case “did not attempt to quantify the causes of the gap and was in any event inadmissible under the *Daubert* test.” Id. at 537. As he had in Sheehan, Judge Posner stated that the Daubert analysis requires “that the methods used by the expert to derive his opinion satisfy the standards for scientific methodology that his profession would require of his out-of-court research.” Id. (citing Sheehan, 104 F.3d at 942; Braun v. Lorillard Inc., 84 F.3d 230, 235 (7th Cir. 1996); Raynor v. Merrell Pharm., Inc., 104 F.3d 1371, 1375 (D.C. Cir. 1997)). After pointing out that the study did not measure poverty, the educational attainments of a student's parents or the extent of parents' involvement in their upbringing, Judge Posner concluded:

A statistical study is not inadmissible merely because it is unable to exclude all possible causal factors other than the one of interest. But a statistical study that fails to correct for salient explanatory variables, or even to make the most elementary comparisons, has no value as a causal explanation and is therefore inadmissible in a federal court.

Id. at 357-58.

In both cases Judge Posner deemed inadmissible statistical evidence where the experts did not control for variables other than age (in Sheehan) and race (in People Who Care) on the basis that the expert did not exercise the degree of care that that a professional should use in his scientific work outside the litigation context. Neumark did not control for any variables other than race (although it appears the plaintiffs may have asked him not to; during his deposition, Neumark said more than once that he was not asked to do anything more than compare the Infosys data with the relevant labor market and look for patterns). Neumark did not follow the procedure that he said the “large body of research on discrimination in labor markets” followed: “estimating the effects of membership in a protected group, controlling for other factors that might affect the outcome in question.” Dkt. No. 97-1 at 145. Following Judge Posner’s terse reasoning, the court could conclude on this basis that Neumark’s analyses were so unreliable as to be inadmissible. On the other hand, if Neumark’s analyses were otherwise reliable, his inconsistency would seem to go more to the weight the trier of fact should accord his analyses than to their admissibility. The court concludes that, standing alone, Neumark’s failure to follow methods he previously espoused renders his analyses so unreliable that they should not be admitted.

The defendants also assert that Neumark’s failure to consider alternative explanations for the disparities he observed—regardless of whether he used the standard of care he would have used outside the litigation context—renders his opinions so unreliable as to make them inadmissible. This is one of several issues where both parties claim that the Seventh Circuit “consistently” has ruled in favor of their diametrically opposing views; the defendants state in their reply brief that:

Plaintiffs argue that “the Seventh Circuit has *consistently* found that expert testimony regarding a statistical disparity is relevant, even if the expert’s analysis did not include a multiple regression analysis accounting for factors other than race.” Pls.’ Resp. Br. 10 (emphasis added), *citing Adams v. Ameritech Services Inc.*, 231 F.3d 414, 427-28 (7th Cir. 2000); *Mister v. Ill. Central Gulf R.R. Co.*, 832 F.2d 1427, 1431 (7th Cir. 1987). This is precisely backwards: the Seventh Circuit has “consistently” found expert opinions that fail to control for relevant variables entitled to no weight. *People Who Care v. Rockford Bd. of Educ. Sch. Dist. No. 205*, 111 F.3d 528, 537-38 (7th Cir. 1997); *see also Schultz v. Akzo Nobel Paints LLC*, 721 F.3d 426, 433 (7th Cir. 2013) (noting an expert opinion that “[d]oes not rule in any causes of [the basis for the claim], nor . . . rule out anything” is not probative of anything and unreliable); *Sheehan v. Daily Racing Form, Inc.*, 104 F.3d 940, 942 (7th Cir. 1997) (holding expert report that failed “to make any adjustment for variables bearing on the decision whether to discharge or retain a person on the list other than [alleged discrimination]” was not “admissible under the standard of *Daubert*”).

Dkt. No. 128 at 6-7.

Again, the parties talk past each other. The *plaintiffs* maintain that the Seventh Circuit has found expert analyses “relevant” even if the expert did not consider variables other than the salient factor. Evidence is “relevant” if it has “any tendency to make a fact more or less probable than it would be without the evidence.” Fed. R. Evid. 401. Under Daubert, reliability is a prerequisite for

relevance; to decide whether expert testimony will “assist the trier of fact to understand or determine a fact in issue,” the court must make a “preliminary assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue.” Daubert, 509 U.S. at 592-93.

The *defendants* maintain that the Seventh Circuit has held that expert analyses that do not consider variables other than the salient factor are “entitled to no weight.” Dkt. No. 128 at 6. Under Fed. R. Evid. 403, a court may exclude relevant evidence “if its probative value is substantially outweighed” by a risk of unfair prejudice, confusion or cumulativeness. Even if the expert’s methodology was scientifically valid and could properly be applied to the facts at issue, a judge may exclude it for the reasons listed in Rule 403. “Expert evidence can be both powerful and quite misleading because of the difficulty in evaluating it. Because of this risk, the judge in weighing possible prejudice against probative force under Rule 403 . . . exercises more control over experts than over lay witnesses.” Daubert, 509 U.S. at 595 (quoting J. Weinstein, RULE 702 OF THE FEDERAL RULES OF EVIDENCE IS SOUND; IT SHOULD NOT BE AMENDED, 138 F.R.D. 631, 632 (1991)).

Both arguments beg the question—have the Supreme Court and the Seventh Circuit “consistently” held that statistical evidence resulting from analyses that do not consider variables other than the salient characteristic are *per se* unreliable under Daubert—particularly in the context of pattern-or-practice discrimination cases? The answer is no.

In Bazemore v. Friday, the plaintiffs had alleged that a state agricultural extension service had engaged in a pattern or practice of employment discrimination based on race, and after a “lengthy trial,” the district court had ruled in favor of the extension service. Bazemore v. Friday, 478 U.S. 385, 386 (1986) (per curiam). At trial, the plaintiffs had “relied heavily on multiple regression analyses designed to demonstrate that blacks were paid less than similarly situated whites.” Id. at 398 (Brennan, J. concurring in part). Some of the regressions “used four independent variables—race, education, tenure, and job title.” Id. Both the trial court and the appeals court had excluded those analyses because they did not include “all measurable variables thought to have an effect on salary level.” Id. at 400 (citing the Court of Appeals). The Supreme Court disagreed:

The Court of Appeals erred in stating that petitioners’ regression analyses were “unacceptable as evidence of discrimination,” because they did not include “all measurable variables thought to have an effect on salary level.” The court’s view of the evidentiary value of the regression analyses was plainly incorrect. While the omission of variables from a regression analysis may render the analysis less probative than it otherwise might be, it can hardly be said, absent some other infirmity, that an analysis which accounts for the major factors “must be considered unacceptable as evidence of discrimination.” . . . Normally, failure to include variables will affect the analysis’ probativeness, not its admissibility.

Id. In a footnote, the Court said, “There may, of course, be some regressions so incomplete as to be inadmissible as irrelevant; but such was clearly not the case here.” Id. at n.10. Bazemore did not address the reliability of statistical analyses that did not include *any* variables other than the salient factor, but it

left open the possibility that such an analysis might be inadmissible “as irrelevant.” Id.

The following year, the Seventh Circuit addressed an appeal in a class action alleging that the defendant railroad had violated Title VII by discriminating in hiring based on race. In Mister, 832 F.2d at 1429, the district court had found in favor of the defendant, concluding that “the class had established a *prima facie* case of discrimination under both disparate treatment and disparate impact approaches, but that the [defendant] had demonstrated that a neutral rule—the desire to hire laborers who lived close to work—not only accounted for the disparity but also was supported by business necessity (a requirement in the disparate impact portion of the case).” Id. The district court had based its conclusion that the class had proved its *prima facie* case on several kinds of evidence, including expert witness testimony that there was less than one chance in a million that the disparity in hiring between white and Black applicants was consistent with race-neutral hiring. Id. On appeal, the defendant challenged the statistical evidence “on two grounds: bad data and inaccurate assumptions about the labor market.” Id. at 1430. The court quickly disposed of the bad data argument, finding that “[t]he plaintiffs’ expert used the best data available; that the data were not better is [the defendant’s] fault, and we agree with the district court that the data at hand were good enough even though imperfect.” Id.

As to the expert’s inaccurate assumptions about the labor market, the court concluded that the plaintiff had “made out a presumptive case of

disparate treatment.” Id. at 1431. The plaintiffs characterize this as support for their contention that “expert testimony regarding a statistical disparity is relevant, even if the expert’s analysis did not include a multiple regression analysis accounting for factors other than race.” Dkt. No. 123 at 14. It is not clear that Mister supports such a broad and general proposition. In fact, the Mister court criticized the experts’ failure to consider variables other than race:

The plaintiffs’ expert used no independent variables other than race. He assumed, in other words, that all applicants are identical in every respect except race. This is not necessarily true. For laborers’ jobs, other variables matter—physical condition (for which age and a weight/height ratio may be proxies) and employment history (has the person been fired from other jobs or convicted of job-related offenses?) are important to any employer. The omission of these variables weakens the plaintiffs’ case by leaving open the possibility that important, non-racial variables account for the hiring decisions.

Id. at 1431.

Despite these criticisms, the Seventh Circuit observed that the defendant had not suggested that any of these variables accounted for its hiring patterns—the defendant’s expert had “replicated the findings of the plaintiffs’ expert,” and the defendant’s only explanation for the “startling disparity” in hiring rates was “distance from work.” Id. Given that, the court said, “We are not about to discount the plaintiffs’ statistical work on grounds that the employer, with the best access to data, chose not to raise,” and it agreed with the district court that the plaintiffs had made out their *prima facie* case. Id.

The Mister court did not cite Daubert. It did not refer to Rule 702 or the Daubert factors when considering the expert’s methods. It did not state that the expert’s opinions were “relevant.” Despite the deficiencies it found with the

statistical evidence, the Seventh Circuit refused to disturb the district court's conclusion that the plaintiffs' statistical evidence had proven the plaintiff's *prima facie* case at trial given the defendant's failure to rebut it. But it stated the obvious—that an expert's failure to consider variables other than the salient characteristic “weakens” the plaintiffs' case (in other words, entitles the statistical evidence to less weight).

In Mozee, 940 F.2d at 1042 the plaintiffs produced at trial “statistical evidence of disparate promotion discipline practices . . . which was probative of . . . a pattern or practice of disparate treatment.” Like Mister, Mozee does not mention Rule 702 or Daubert and the court did not analyze the expert's methods or opinions under either. Like the Mister court, the Mozee court considered the statistical evidence in the post-trial context; the defendant argued on appeal that the “pool” against which its workforce had been compared did not account for the qualifications for the particular position to which the plaintiffs alleged they'd been unlawfully denied promotion. Id. at 1045. The Seventh Circuit said that

“[n]ormally, failure to include variables will affect the analysis' probativeness, not its admissibility.” *Bazemore v. Friday*, 478 U.S. 385, 400 . . . (1986). We have ourselves understood this principle to require that plaintiffs eliminate “the most common non-discriminatory reasons” for any suggested disparity. *Coates [v. Johnson & Johnson]*, 756 F.2d [524,] at 541 [(7th Cir. 1985)] (quoting *Texas Dep't of Community Affairs v. Burdine*, 450 U.S. 248, 254 . . . (1981)).

Id. at 1045. Mozee again stated the obvious—that failure to account for variables other than the salient characteristic reduces the probative value of



the statistical evidence (entitles it to less weight) and that it is, if not a requirement, then a “best practice” that a plaintiff’s expert account for at least the most common non-discriminatory variables.

In Sheehan and People Who Care, Judge Posner *did* analyze the statistical evidence under the Daubert standard. He equated a statistician’s failure to account for variables other than the salient characteristic with a “failure to exercise the degree of care that a statistician would use in his scientific work, outside of the context of litigation,” Sheehan, 104 F.3d at 942, which he concluded rendered the analyses inadmissible.

That brings the court to Adams—a pattern-or-practice case of alleged age discrimination that each party claims supports its position. In Adams, the Seventh Circuit reversed the district court’s grant of summary judgment, concluding that “the plaintiffs presented enough evidence to withstand the defendants’ motions.” Adams, 231 F.3d at 417.

The Adams court applied the Daubert factors to the statistical evidence the plaintiffs had proffered. Id. at 423. The defendants had “questioned whether statistical evidence as a whole can ever be useful in a case alleging disparate treatment or a discriminatory pattern or practice, as opposed to a disparate impact case.” Id. The court answered that question in the affirmative, explaining that “statistical evidence can be very useful to prove discrimination in either or both of those two kinds of cases, but it will likely not be sufficient in itself.” Id. (citation omitted). It observed that the Supreme Court had used such evidence in Hazelwood, approving the use of statistics “to help show that

the school district was discriminating on the basis of race in its faulty hiring decisions,” but underscoring “the importance of looking to the proper ‘community’ or group when making statistical comparisons.” Id. (citing Hazelwood, 433 U.S. at 307, 310-12). The Adams court noted that in Hazelwood, the Court had held that the defendant “was entitled to an opportunity to rebut any inference of discrimination raised by the plaintiffs’ statistical showing.” Id. (citing Hazelwood, 433 U.S. at 309-10). It also quoted the Bazemore Court’s holding that “[n]ormally, failure to include variables will affect the analysis’ probativeness, not its admissibility,” and recounted the Bazemore Court’s reminder that courts must evaluate statistical evidence in light of the remaining record evidence. Id. (quoting Bazemore, 478 U.S. at 400).

The Adams court then turned to its own decisions considering the use of statistical evidence, particularly Radue. Id. The court explained that in Radue, it had upheld a grant of summary judgment in favor of the employer in an ADEA case “where the employee’s *prima facie* case was primarily composed of statistics that showed that older employees were treated less favorably than younger employees in various RIFs [reductions in force] the employer had carried out.” Id. The court explained, however, that

those statistics were flawed in a number of ways, and the plaintiff had little else with which to support his case. For one thing, the statistics looked at a completely different part of the company from the one in which the plaintiff worked and they involved an earlier RIF, not the one in which the plaintiff lost his job. [Radue, 219 F.3d] at 616. For another, the statistics failed to address the essence of the plaintiff’s claim: he did not allege that the RIF was age-based; he claimed instead that the transfers awarded in the wake of the RIFs were given preferentially to younger employees. *Id.* Finally, the court noted that the statistics standing alone could not prove causation;

they could only show a relation between the employer's decisions and the affected employees' traits. *Id.* See also *Tagatz v. Marquette Univ.*, 861 F.2d 1040, 1044 (7th Cir. 1988) ("Correlation is not causation.").

Id. at 423-24.

The Adams court also discussed Mister, recounting the process of conducting a study in employment discrimination cases (defining the relevant labor market, finding the null hypothesis results, then comparing the defendant's results to the null hypothesis and determining the extent of any disparities); the court stated that "[t]wo standard deviations is normally enough to show that it is extremely unlikely (that is, there is less than a 5% probability) that the disparity is due to chance, giving rise to a reasonable inference that the hiring was not race-neutral; the more standard deviations away, the less likely the factor in question played no role in the decisionmaking process." Id. at 424 (citations omitted).

The court observed that "*Mister* . . . noted the importance of making sure that any testing adequately accounts for the real variables that the employer took into account." Id. The court explained that "[t]he plaintiffs' evidence [*in Mister*] was wanting because it assumed that the job applicants were identical in every respect except for race, even though other factors like physical condition, employment history, or other non-racial variables might have entered into the employer's calculus." Id. (citing Mister, 832 F.2d at 1431). The Seventh Circuit qualified, however, that

it was [the defendant] that had the responsibility of offering alternative explanations. The only one it actually advanced was that it was concerned about the distance its employees had to travel to

get to the job site; for reasons pointed out in the opinion, this was singularly unpersuasive. (It is also important to note that *Mister* reached this court after a full trial on the merits; nothing in the opinion suggests that the plaintiffs' statistical showing was so weak that summary judgment for the defendant would have been appropriate.)

Id. at 424-25.

Given this history, the Adams court rejected the defendants' argument that statistical evidence never could be useful in a case alleging disparate treatment discrimination, indicating that it was "hard pressed" to understand that position given the Supreme Court's rulings. Id. at 424. Having laid that foundation, the court turned to the issue before this court: "whether, under the *Daubert* standards, the statistical evidence plaintiffs have offered . . . (coupled with the other evidence they presented) was sound enough methodologically (*i.e.*, reliable enough) and relevant, such that the district court should have taken it into account in evaluating their claim." Id. at 425. It explained that its task was to determine "whether the criticisms of the [expert's] reports and the plaintiffs' other statistical evidence affected the admissibility of those materials, or only, as the Supreme Court put it in *Bazemore*, their 'probativeness' or weight." Id.

The court seems to have answered that question by concluding that the defendants' criticisms went more to the "probativeness," or weight, of the expert's opinions. The Adams court observed that its task in conducting a *de novo* review of a summary judgment ruling was not to determine which party's expert report was more persuasive, but whether, "taking the facts in the light

most favorable to the plaintiffs, a trier of fact should be permitted to make that choice.” Id. Buried in the parties’ debate over the relative merits of each of their experts were some factual issues, including the question of how the statistical expert should have aggregated, or disaggregated, the defendants’ workforce. Id. The court explained what the expert had done, then noted what he had *not* done: “run a multiple-regression analysis that would have isolated the relevance of age as a factor in the companies’ decisions.” Id. The court stated, “While this omission strikes us as odd, we are not prepared to hold as a matter of law that nothing but regression analyses can produce evidence that passes the *Daubert* and *Kumho Tire* thresholds.” Id. The court explained:

Statisticians might have good reasons to look at data in different ways. (For example, as additional variables are introduced into a regression, the less likely it is that any of them will be statistically significant, a fact that causes its own problems.) We thus evaluate here what [the expert] did, rather than hypothetical tests that he or another expert might have done.

Id.

The court detailed how the expert had conducted his study and how he had analyzed his findings. Id. at 425-27. It also explained why the district court had declared the expert’s reports inadmissible:

(1) the underlying information about the RIF programs was not reliable; (2) the reports only showed that the differences in treatment between the over and under 40 aged individuals was not due to chance, but they did not affirmatively indicate what caused that difference; (3) the analysis did not take into account or control for other non-age related variables; (4) [the expert] relied on the plaintiffs’ description of the RIF and did not himself become familiar with the procedures used; and (5) the jury would find the reports so confusing that they should be excluded under Rule 403.

Id. at 427.

Unlike the district court, the Seventh Circuit saw no problem with the expert relying on data that ultimately had come from the defendants, or with how he familiarized himself with the procedures the defendants had used to reduce the workforce. Id. It characterized as “more serious,” however, the defendants’ second and third objections, describing as their “theme” that the expert’s analysis “standing alone, was not enough to show that age was the reason why [the defendants] took the actions that they did.” Id. The court agreed; it said that “the statistical analyses were enough to rule out chance, but the real reason for the decisions may have been age or it may have been some other factor or factors positively correlated with both advancing age and the likelihood of termination.” Id.

But, the court stated, “ruling out chance was an important step in the plaintiffs’ proof, even if it was not a single leap from the starting line to the finish line.” Id. The court said that if the statistical analysis had been “all the plaintiffs had introduced, we would agree with the district court that the record would have supported summary judgment against them.” Id. But it concluded that statistical analyses were not the only evidence the plaintiffs had produced; “in our view the other items of evidence, if believed by a jury, could have done the rest of the job: that is, it could have ruled out factors other than age.” Id. at 427-28.

The court concluded by reiterating that it was reviewing a grant of summary judgment; it stated, “[w]e hold only that [the statistical] evidence met the standards of admissibility set by the Federal Rules of Evidence and thus

should have been counted on plaintiffs' side for summary judgment purposes.”  
Id. at 428.

This review indicates that the statements in Sheehan and People Who Care that an expert's failure to consider variables other than the salient characteristic renders his opinions inadmissible are outliers. There are some common themes in the cases: statistics play an important role in a plaintiffs' ability to make a *prima facie* case of pattern-or-practice discrimination; experts should carefully formulate their initial hypotheses; experts should consider variables other than the salient characteristic and their failure to do so weakens and renders less probative their conclusions. But particularly when plaintiffs have other evidence (such as individual testimony about personal experiences of discrimination), the weight of decision seems to treat failure to consider other variables as a factor going to the weight of the statistical evidence and not as a factor that renders the evidence inadmissible. The court concludes that Neumark's failure to follow what the Seventh Circuit appears to deem “best practices” by failing to consider variables other than the salient factor of race does not render his opinions so unreliable that the statistical evidence is irrelevant on that basis alone. The court concludes that if the only flaw in Neumark's methodology had been his failure to account for variables other than race, the evidence would have been relevant to the plaintiffs' allegations that Infosys discriminated under Rule 401's “liberal,” basic standard of relevance, Daubert, 509 U.S. at 587, and the defendants' criticisms

would go to the probative value, or weight, that the trier of fact should give that evidence.<sup>16</sup>

Before leaving this topic, the court addresses the plaintiffs' argument that Neumark was not required to consider alternative variables because the burden is on Infosys, not the plaintiffs, to prove a race-neutral reason for the disparities the statistics revealed, and that Infosys "failed to carry its burden." Dkt. No. 123 at 16. As the defendants have observed, this argument confuses the plaintiffs' burden of proof at trial under the Teamsters framework with the reliability requirement of Daubert. The plaintiffs are correct that under Teamsters, if the plaintiffs prove their *prima facie* case of a widespread pattern or practice of discrimination, the burden will shift to Infosys to demonstrate race-neutral explanations for the disparities. But to prove that *prima facie* case, the plaintiffs must present competent statistical evidence that proves by a preponderance of the evidence that there was a systemwide pattern or practice of discrimination. Because the only evidence upon which the plaintiffs rely to carry that burden is Neumark's statistical analyses, those analyses must pass scrutiny under Daubert—they must be based on reliable methodology. The

---

<sup>16</sup> The plaintiffs also argue that Neumark's general methodology was both relevant and reliable because he followed the steps the Seventh Circuit has outlined in Adams and Mister: he defined the relevant labor market, then asked what the result would be for the salient variable absent discrimination, then compared that result with the defendants' data and computed the standard deviations from the null hypothesis. Dkt. No. 123 at 13-14 (quoting Adams, 231 F.3d at 424). That methodology is reliable only if the expert's determination of the salient variable is reliable. Neumark's was not.



question of whether the expert considered alternative variables is a component in determining whether he utilized reliable methodology.

b. Race Identification Methodology

While Neumark's failure to consider alternate variables does not render his analyses and opinions unreliable, Neumark's race identification methodology bears none of the hallmarks of reliability. "Ordinarily, a key question to be answered in determining whether a theory or technique is scientific knowledge that will assist the trier of fact will be whether it can be (and has been) tested." Daubert, 509 U.S. at 593. Neumark's method of name-matching and exclusion to determine whether someone is Indian or South Asian has not been tested. Perhaps one could come up with an empirical test to determine whether Neumark's method was accurate, but the plaintiffs have not argued as much nor identified such a test. Neumark's methodology has not "been subjected to peer review and publication." Id. There is no evidence of "the known or potential rate of error" of his methodology, or of "the existence and maintenance of standards controlling" the methodology. Id. at 594. There is no evidence that the methodology is generally accepted in a "relevant scientific community." Id. (quoting United States v. Downing, 753 F.2d 1224, 1238 (3d Cir. 1985)). The race recognition methodology did not arise naturally and directly out of research Neumark conducted independent of the litigation, see Daubert v. Merrell Dow Pharm., 43 F.3d 1311, 1317 (9th Cir. 1995); Neumark and the plaintiffs admit that Neumark devised the methodology for this case. Neumark has not even extrapolated from an "accepted premise to an

unfounded conclusion,” Committee Notes, 2000 Amendment to Fed. R. Evid. 702; see Gen. Elec. Co. v. Joiner, 522 U.S. 136, 146 (1997); his “premise” for determining which names are likely to be Indian or South Asian versus “Western” is not founded.

The plaintiffs try to address this problem in several ways. They argued in their brief and at the hearing that there is no “perfect” way to determine someone’s race when the data do not include that information.<sup>17</sup> They are correct that the issue is not whether Neumark’s method was perfect, but the fact that the method Neumark used need not be perfect does not address whether the method he used was *reliable*. The plaintiffs asserted in their brief that Neumark’s race-identification methodology was “reasonably accurate,” but provided no support for that assertion. Dkt. No. 123 at 2. The plaintiffs argued in the brief that Neumark’s findings are consistent with the PeopleFluent “analyses.” Id. Setting aside the defendants’ extensive arguments that the “computations and affirmative action goals assembled by PeopleFluent” are not evidence of discrimination, are not admissible and do not constitute expert

---

<sup>17</sup> It appears that there are experts in “surname analysis” and that there are methodologies for making the kind of determination Neumark tried to make here. See, e.g. U.S. v. Johnson, 122 F. Supp. 3d 272, 338-39 (M.D.N.C. 2015); NAACP, Spring Valley Branch v. E. Ramapo Cent. Sch. Dist., 462 F. Supp. 3d 368, 382 (S.D.N.Y. 2020) (discussing “Bayesian Improved Surname Geocoding (BISG),” a “methodology that uses individual-level data, including a voter’s surname, geographic location, and the racial composition of the voter’s census tract or block to generate the probability that an individual belongs to a particular group where self-reported information is not available”); M. Elliott, A. Fremont, P. Morrison, P. Pantoja and N. Lurie, A NEW METHOD FOR ESTIMATING RACE/ETHNICITY AND ASSOCIATED DISPARITIES WHERE ADMINISTRATIVE RECORDS LACK SELF-REPORTED RACE/ETHNICITY, Health Serv. Research Oct. 2008, available at <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2653886/>.

opinions, *dkt. no. 125* at 3-9, the fact that something or someone else reached a similar conclusion to Neumark does not mean that the method he used to reach his conclusions was reliable. The court's focus "must be solely on principles and methodology, not on the conclusions that they generate." Daubert, 509 U.S. at 595. The plaintiffs argue in their brief that Infosys "substantially overstates the effect of the limited number of misclassifications." *Dkt. No. 123* at 23. It is not clear how the plaintiffs can make this assertion, particularly regarding Neumark's classification of applicants for whom there is no nationality or racial data.

Finally, the plaintiffs argued at the motion hearing that "name analysis" had "been done" in Teamsters, asserting that "[i]n *Teamsters*, they use surnames to identify individuals of Hispanic d[escent]." *Dkt. No. 204* at 48, lines 11-13. As the court tried—unsuccessfully—to point out at the hearing, this is an inaccurate characterization. The government alleged in Teamsters that the defendants had engaged in a pattern or practice of discrimination against "Spanish-surnamed Americans," Teamsters, 431 U.S. at 328—people who had "Spanish" last names. As the court stated at the hearing, a person with a "Spanish" surname is not necessarily a person of "Hispanic descent." More to the point, the plaintiff in Teamsters did not allege that the employer had discriminated against people with Spanish surnames because of their race (though it did allege that the employer had discriminated against Black people on that basis), and the Teamsters Court was not asked to resolve a dispute

about how the plaintiff's expert decided whether a person had a "Spanish" surname or whether a particular surname was "Spanish."

Given the plaintiffs' insistence that Neumark's race identification methodology was reliable, one is tempted to analyze weaknesses in the base assumptions underlying that methodology—that every person born in India and South Asia is racially Indian or South Asian; that surnames always correlate to race and never to marriage or adoption; that someone not born in India or not of South Asian race may have a surname that is common in India or South Asia that does not appear on the Infosys employee roster. If Neumark had used reliable methodology in reaching his conclusions, those weaknesses might have gone to the weight a factfinder would give the conclusions, but he did not use a reliable methodology and the plaintiffs have not come close to demonstrating that he did.

Interestingly, the defendants argue that Neumark's race classification methodology "in his applicant studies . . . is irrational, arbitrary and unscientific," *dkt. no. 98 at 29*, implying that the methodology was reliable for Neumark's other studies. It is not clear why Neumark's methodology would be flawed only with regard to the applicant studies. Neumark used his matching and excluding methodology to determine race—specifically, South Asian descent—and in the promotion and termination studies he did so not only for applicants but for incumbent employees. Because the Infosys data did not contain self-reported information about whether an employee considered himself or herself to be South Asian, Neumark was required to use his

methodology—created for this litigation, not tested, not peer-reviewed, with no known history of error rates—to determine which incumbent Infosys *employees* were “South Asian,” not just which *applicants* were likely to be “South Asian.”

The court cannot conclude that Neumark’s race identification methodology was reliable. He used that methodology to determine the race of incumbent employees and applicants and used the results to conduct the comparative studies that resulted in the disparities he identified in his report and upon which the plaintiffs rely.

### 3. *Relevance*

Having determined that Neumark was not qualified to determine the race of Infosys employees or applicants by surnames and that the methodology he used to do so was unreliable, the court must conclude that his analyses are not relevant under Daubert. The plaintiffs repeatedly have asserted that statistics showing disparities in hiring, promotion and termination are relevant to their *prima facie* case that Infosys had a pattern or practice of discrimination against non-Indian and non-South Asian employees and applicants. That is true only if those statistics would “assist the trier of fact to understand the evidence or to determine a fact in issue.” Rule 702(a). “Rule 702’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.” Daubert, 509 U.S. at 591-92. Neumark’s statistical analyses are based on racial identifications that he was not qualified to make and that he made using unreliable methodology. Because they lack that “valid scientific connection” to the pertinent inquiry—were the people he

classified as Indian or South Asian actually Indian or South Asian?—

Neumark's analyses are not relevant.

The court is mindful that proving a *prima facie* case of a pattern or practice of discrimination in a company the size of the defendants' is expensive. Plaintiffs may not be in a position to retain a "name analysis," or "race identification," expert *and* a statistical expert *and* a labor economist, and some cases may not call for it. Dr. Neumark is an experienced scholar and prolific writer whose credentials in the field of labor economics the defendants have not challenged. The named plaintiffs have described incidents and experiences which range from unpleasant and uncomfortable to frightening. The court's conclusion that Neumark's report must be excluded because it does not pass muster under Daubert is not a condemnation of the plaintiffs' case or of Neumark, nor is it a prediction of how a trier of fact might ultimately decide the plaintiffs' claims.

The court will grant the defendants' motion to exclude Neumark's expert opinions; it deems both the September 2016 and February 2017 reports and the analyses and opinions contained in them inadmissible.

## **VII. The Remaining Motions**

The plaintiffs' motion for partial summary judgment is based entirely on Neumark's analyses. Dkt. No. 86. The court will deny that motion without prejudice.

The plaintiffs' motion for class certification explicitly relies on Neumark's analysis to show numerosity and predominance, and it likely ought to have

done so for commonality. Dkt. No. 88. The court will deny that motion without prejudice.

The court's extreme delay in ruling on the motion to exclude Neumark's opinions caused the parties to seek a hearing for the purpose of determining whether they could do anything to assist the court in making its decision. Dkt. No. 202. This ruling renders that motion moot.

The plaintiffs' brief in opposition to the defendants' motion for summary judgment (Dkt. No. 93) and its combined additional proposed findings of fact and response to the defendants' proposed findings of fact (Dkt. No. 94) rely on and are replete with reference to Neumark's analyses.

Given the court's delay in ruling on the Daubert motion and the impact this ruling has on the plaintiffs' case, the court will give the parties an opportunity to digest this decision. The court will schedule a status conference to discuss with the parties proposed next steps.

### **VIII. Conclusion**

The court **GRANTS** the defendants' motion to exclude the expert opinions of David Neumark. Dkt. No. 97.

The court **DENIES WITHOUT PREJUDICE** the plaintiffs' motion for partial summary judgment. Dkt. No. 86.


The court **DENIES WITHOUT PREJUDICE** the plaintiffs' motion for class certification. Dkt. No. 88.

The court **DENIES AS MOOT** the parties' joint motion for status conference. Dkt. No. 202.

The court **ORDERS** that the parties must appear for a telephonic status conference on **November 10, 2022 at 1:30 PM**. The parties are to appear by calling the court's conference line at 888-557-8511 and entering access code 4893665#.

Dated in Milwaukee, Wisconsin this 14th day of September, 2022.

**BY THE COURT:**



---

**HON. PAMELA PEPPER**  
**Chief United States District Judge**